THE CHICAGO MEDICAL SCHOOL
QUARTERLY

VOLUME 20 SEPTEMBER, 1959 NUMBER 3

HAZARDS AND SAFEGUARDS IN HEMOTHERAPY *

KURT STERN, M.D.**

"In numbers there is safety," is a rule which we commonly accept for many aspects of daily life. However, when applied to blood transfusions, this saying may well have to be converted into its opposite: "In numbers there is danger!" The present scope of hemotherapy is only partly indicated by results of surveys[1] according to which an estimated 5 million pints of blood are collected annually in the United States and used for more than 4,500,000 transfusions given to 2,000,000 patients.

Such figures might tempt one to propose a revision of the classic Harveyan description of blood circulation referring to transfusion therapy as an extracorporeal shunt by means of which significant amounts of blood are diverted from the vascular compartment of one group of the population (blood donors) into sundry forms of glass and plastic containers before reentering the system of another group of persons (recipients).

While there is no intent to minimize the value of this "extracorporeal blood circulation" as an indispensible prerequisite for many achievements of modern medicine, the ready availability of blood and relative ease of administration of transfusions make it imperative to review periodically hazards connected with hemotherapy and to point out safeguards to be observed in order to avoid serious damage to patients from this form of treatment.

Since each blood transfusion even under optimal conditions is associated with certain inherent risks, it is a logical conclusion that blood must never be given unless there is an adequate indication which justifies taking the associated calculated risks. In other words, there is no room for "cosmetic" hemotherapy or for giving blood as a form of "moral support" for the patient and his family, or to rely on the mystic concept of a "tonic" effect of blood.

It is quite likely that a sizable percentage of single units of blood administered to a particular patient could, on closer examination, not be justified as valid form of hemotherapy. Indications for blood transfusions have been analyzed in detail only infrequently. In one such study[2], valid indications were found to be lacking in approximately 13 per cent of 290 transfusions studied.

In the following presentation, an attempt will be made to survey the most important complications of blood transfusions. The emphasis will be placed on clinical aspects rather than on collection, preservation, and selection of blood which must be considered as the main responsibility of the blood bank staff.

* Departments of Pathology, The Chicago Medical School; Mount Sinai Hospital; and the Blood Center, Mount Sinai Medical Research Foundation.

** Associate Professor of Pathology, The Chicago Medical School; Director, Mount Sinai Blood Center.

I. TRANSFUSION REACTIONS

By definition, these are untoward effects occurring either during hemotherapy or shortly thereafter for which a more or less definite causal relationship to the transfusion can be established.

1. **Hemolytic transfusion reactions.** This most serious form probably accounts for the highest percentage of fatalities resulting from blood transfusions. With rare exceptions, such reactions are today preventable, provided that blood has been selected by means of adequate pretransfusion or cross-matching tests which can insure that the recipient does not possess any antibody for the red cells of the prospective donor and that, vice versa, the prospective donor is free of antibodies that could agglutinate and thus damage the red cells of the recipient. However, tragic experiences have time and again confirmed the regrettable fact that most hemolytic transfusion reactions are not the result of technical inadequacies or incompetence of blook bank personnel, but rather can be traced to human errors, such as mistakes in identity of donor, recipient, or both. This reemphasizes the paramount need for proper organization of transfusion services including suitable instruction of all medical and nursing personnel who participate in administration of blood transfusions.

A relatively small number of hemolytic transfusion reactions can subsequently be proven to have been caused by destruction of transfused blood by hemoantibodies in the recipient which prior to the transfusion were not detectable because their quantity was below the threshold of sensitivity of technical methods currently available. Fortunately such reactions are, as a rule, mild and if the recipient is properly managed, do not entail serious consequences. It is, however, imperative that each transfusion reaction be adequately studied and be completely elucidated before any additional blood is given. Only in this way is it possible to avoid more serious consequences that may result from giving additional blood containing the offending antigen. Hence, adequate procedures and methods for investigation of each transfusion reaction are of utmost importance[3]. Actually, differential diagnosis of other forms of transfusion reactions to be discussed further on is important mainly inasmuch as this eliminates the possibility that it was a hemolytic reaction.

Since clinical symptoms by themselves are notoriously unreliable for such differential diagnosis, one must subject each transfusion reaction to a careful investigation which is capable of ruling out the possibility of a hemolytic episode before one can accept the diagnosis of another type of reaction.

Some manifestations of hemolysis after transfusions may not be caused by antigen-antibody reactions but by imperfect storage of the blood or by administration of unsuitable intravenous fluids or medications together with the blood. Special attention is called to the fact that five per cent glucose in water, although supposedly isotonic, must never be given through the same administration set as blood, since this may bring about hemolysis[4, 5]. Only physiologic saline or 5 per cent glucose in physiologic saline are safe for simultaneous administration with blood using the same tubing while all other intravenous fluids should be given by means of a different administration set. By the same token, direct admixture of medications to units of blood should be avoided whenever possible, both because of possible untoward interactions of the drug with blood components as well as because of the danger of introducing bacterial contamination.

2. **Pyrogenic or febrile reactions.** These side effects of transfusions which become manifest as chills and/or elevation of temperature, used to be exceedingly common when it was difficult to keep reusable blood transfusion equipment entirely free of bacterial pyrogens, e.g., proteinaceous material of bacterial origin that became attached to rubber tubing. With the introduction and wide-spread use of disposable equipment, febrile transfusion reactions have been markedly reduced but not completely eliminated. In several large series with adequate follow-up they occurred in from 0.7 to 2 per cent of all transfusions.

Within recent years a new important etiologic factor has been demonstrated

as responsible for some of these pyrogenic reactions, namely, development in the recipient of leukagglutinins (isoantibodies for white cells)[6-9]. The incidence of such leukagglutinins was found to rise with the number of transfusions given to individual recipients. In experiments on volunteers such leukagglutinins could be produced when they were transfused seven times with blood of the same donors, and the resulting antibodies exhibited a specificity indicating that antigenic differences exist between leukocytes of different persons[10].

In addition to the demonstration *in vitro* of white cell antibodies, their etiologic importance may be inferred from the fact that patients who had undergone severe pyrogenic transfusion reactions repeatedly were found to tolerate without any ill effect transfusion of blood prepared in such a manner as to be virtually free of white cells[6,8,9].

While pyrogenic reactions, on the whole, rarely present serious danger to the recipient, they are frequently the cause of considerable discomfort and hence make desirable methods of prevention or treatment. Some authors[11] have recommended the use of antihistaminics for prevention of febrile reactions, but others have found this method ineffective[12,13]. On the other hand, treatment of febrile reactions appears to have been successfully accomplished by means of opium alkaloids[14] although it will be rarely necessary to resort to such drastic measures. On the other hand, an important factor capable of reducing pyrogenic reactions is represented by control of the speed of administration. There is ample evidence that rapidly administered blood increases the incidence of pyrogenic reactions, a phenomenon which reflects the fact that a minimal dose of pyrogen must be administered during a given period of time in order to produce the biologic effect.

3. **Allergic reactions.** These relatively harmless complications of transfusions are most commonly urticarial lesions of skin or mucous membranes, or may present themselves merely as itching. On rare occasions, they cause serious manifestations such as laryngeal edema. In most instances it is impossible to implicate a specific allergen as responsible for the reactions.

In order to reduce or eliminate certain allergic reactions, it is advisable not to accept as donors persons exhibiting active allergic manifestations (urticaria, hay fever). An older concept that allergic reactions may be more common after transfusion of fatty (lipemic) blood has not been corroborated by more recent studies[15]. If treatment is necessary, administration of suitable antihistaminics, as a rule, will accomplish the desired purpose. Only in the face of serious manifestations, such as laryngeal edema, is it necessary to use intravenous calcium gluconate and/or adrenalin.

Several authors[11-13,16,17] have proposed to add antihistaminics to bottles of blood and reported that such measures reduced incidence of allergic reactions to less than one-half of one per cent, or even completely eliminated them. However, it may be questioned as to whether one is justified to add routinely antihistaminic drugs to blood in order to eliminate reactions that may occur possibly in 2 out of 100 transfused patients, while the remaining 98 actually do not need such medication. In other words, it seems to be more reasonable to treat the patient rather than the bottle of blood.

In accord with such a point of view, a committee of the American Academy of Allergy[18] recommended that for prophylaxis of allergic transfusion reactions oral antihistaminic medication should be used rather than addition of the drug to the blood. This is not only safer because some patients may react unfavorably to antihistaminic therapy but also because one cannot exclude the possibility that addition of antihistaminics may have an untoward effect on blood cells. While an antispherocytic effect *in vitro* of phenothiazine derivatives was observed[19-20], further studies showed that *in vivo* survival was not promoted by addition of such compounds[21]. Hence, patients with propensity for developing urticarial manifestations after blood transfusions should be prepared by receiving a suitable antihistaminic 30 minutes prior to transfusion.

In this connection, it should also be re-emphasized that clinical manifesta-

tions, such as chills and fever, or development of wheals do by no means represent proof that one deals with pyrogenic or allergic reactions, since hemolytic transfusion reactions may be associated with similar manifestations[3], a fact that has been corroborated by deliberate administration of small amounts of incompatible blood to experimental subjects[22].

4. **Circulatory overload reactions.** Symptoms of rather serious nature may arise during or following blood transfusions as a result of injudicious quantitative or qualitative use of blood. Signs of acute heart failure are most likely to occur in patients with reduced myocardial reserve, including a large number of aged individuals as well as persons suffering from infectious diseases, such as pneumonia, which place severe strain on the cardiac reserve. Therefore, if such patients require blood, they should be transfused only with small amounts of blood at a time, and if the defect to be corrected concerns mainly red cell mass, they should receive resuspended or packed red cells rather than whole blood, since plasma proteins are the principal cause of expansion of blood volume and overloading of the vascular compartment. Furthermore, the transfusion should be administered at low speed in order to permit gradual adaptation of the vascular compartment.

When disregard of such precautions induces acute heart failure including pulmonary edema, this has therefore also been termed as "speed reaction." Such hazards of overtransfusion must be kept in mind, including the fact that excessive hemotherapy does not lead to arterial hypertension but rather to hypotension[23]. This is well exemplified in a recent report dealing with four patients who, after massive hemorrhage, were overtransfused and showed severe circulatory disturbances leading to death in one case, while the remaining three recovered after institution of appropriate treatment[24].

5. **Embolic reactions.** Complications of this nature should be today merely of historic interest. Early in the development of modern hemotherapy it became apparent that fine fibrin clots may form even under the most favorable conditions of collection and preservation of blood and, therefore, the rule has been generally adopted that mesh filters conforming to specific standards must be interposed between the blood bottle and the administration set. Hence, only as a result of gross negligence, namely, omission of such a filter, would it be possible for blood clots to enter the system of the recipient and cause embolization. The danger of air embolism may be introduced if the administration of blood is speeded up by means of pressure pumps improperly set up or constructed. Obviously such hazards are completely preventable and should not exist in adequately run transfusion services.

6. **Bacterial contamination of blood.** These reactions, fatal in more than 50 per cent of cases, have resulted from administration of blood contaminated with gram-negative bacteria. In one of the first cases reported[25] coliform organisms were isolated from the blood and the clinical manifestations — severe hypotension resulting in a shock-like state and anuria — were found to be attributable to the toxins produced by these organisms.

While in this particular case administration of arterenol was capable of overcoming the hypotensive state with eventual recovery of the patient, in other instances the irreversibility of most of these serious reactions has been stressed[26,27], with fatalities observed after administration of as little as 25 to 50 cc. of blood[27].

Contamination apparently may be introduced at the time of the collection, either due to inadequate sterilization of the skin of the donor or possibly due to the entrance of air contaminants if the blood collecting facilities are not properly taken care of. Certain gram-negative bacilli responsible for these reactions are particularly dangerous because of their ability to grow at low temperatures. This has been extensively substantiated by Pittman[28] and Braude and his associates[29]. Adequate measures for prevention of such reactions demand observance of most rigid asepsis during collection of blood as well as recognition of the fact that any entry into the blood containers except immediately before administration is objectionable as a potential cause

of contamination. Hence, once a certain amount of blood has been administered from a container and the transfusion interrupted for more than a short period, the balance of the blood cannot any more be used safely. Likewise, all measures involving "splitting" of units of blood must be strongly discouraged. If entry into the containers is necessary, such as in preparation of packed or resuspended red cells, this must be carried out with all precautions used for a surgical operation, preferably in a sterile room.

Inspection of the blood before administration may reveal hemolysis or discoloration as potential indicator of contamination, but absence of such signs is no assurance for freedom from this hazard. Some authors have recommended routine addition to blood of tetracycline antibiotics as bacteriostatic agents[29], but such measures are not in general use. Methods for early recognition of reactions due to bacterial contamination of blood have been described in detail and immediate administration of tetracycline antibiotics and continuous intravenous administration of levarterenol have been recommended for their treatment[30]. Obviously, the seriousness of such reactions cannot be overestimated and an ounce of prevention will outweigh by far the many, many pounds of treatment.

7. **Citrate reactions.** Massive transfusion therapy has raised concern about the possible toxic effect of the citrate present in bank blood, since administration of several units of blood within a short period of time might be capable of depressing ionized calcium to levels associated with deleterious effects on the cardiovascular system and hemostasis. Apparently two factors counteract such depressions of calcium: 1) rapid metabolization of citric acid in the liver, 2) rapid mobilization of calcium from the skeletal system as part of physiologic homeostasis.

Extensive investigations[31] have been carried out on chemical changes occurring in 130 patients following transfusion of moderate to large volumes of citrated blood. The highest concentrations of serum citrate were observed in patients with liver disease in whom, therefore, rapid transfusions of large amounts of citrated blood should be avoided. On the other hand, another group of workers failed to find any evidence for citrate intoxication after massive blood replacement, even when patients with damaged liver function were included. Specifically, these authors did not observe any relationship between oozing during surgery and elevations of blood citrate level upon review of 77 patients, 18 of whom received from 16 to 40 units of blood each[32].

Nevertheless, it is mandatory to give special consideration to situations, such as replacement transfusion of infants with hemolytic disease of the new-born, in which it has become established practice to counteract the possible harmful effect of citrate by intravenous injection of calcium gluconate. Similarly it is recognized that in cardiac surgery citrated blood should not be used[33]. A fatality observed after massive transfusion of a patient being operated upon for tetralogy of Fallot was ascribed to citrate intoxication[34].

As discussed later on, for cardiac surgery—especially when requiring extracorporeal circulation — heparinized instead of citrated blood must be employed. Similarly, it has been observed that massive hemotherapy by itself does not cause a drop of ionized calcium, but severe drop of ionized calcium with incoagulability of blood may occur in shock when blood replacement is deficient, and in certain forms of major surgery, especially those involving the thorax[35].

II. PHYSIOPATHOLOGIC EFFECTS OF HEMOTHERAPY

1. **Depression of hematopoiesis.** Animal experimentation has demonstrated that there is at a given time an equilibrium between red cell formation and available red cell mass. In rats made polycythemic by transfusions, a marked reduction in erythropoiesis was observed[36]. Similar events occur also in human anemia; specifically, plasma of children with Cooley's anemia was found to contain a high concentration of an erythropoiesis-stimulating factor which disappeared rapidly after blood transfusions to the patients[37].

From these facts it may be concluded that anemias traceable to specific deficiencies, such as iron, should not be treated with transfusions, unless the clinical condition of the patient precludes waiting for blood replacement to take place as the result of specific treatment. Even then one must realize that transfusions, though rapidly substituting for deficiencies of the red cell mass, depress erythropoietic activity and delay the process of physiologic recovery, a fact corroborated by following the per cent of reticulocytes before and after transfusion. This is especially important in the first 4 to 8 weeks of neonatal life when erythropoietic activity is physiologically at a low level.

2. **Transfusional siderosis.** Deposition of iron in tissues of patients subsequent to transfusion therapy has provoked extensive discussion in the literature. This is not the place to enter into the controversy as to whether or not transfusional siderosis is equivalent to or can be differentiated from endogenous hemochromatosis. Recent studies[38,39] point to striking pathogenetic and pathologic differences between exogenous siderosis and endogenous hemochromatosis. While some patients with transfusional siderosis had received up to eight hundred pints of blood within six to twelve years[38], it is significant that in other patients the condition apparently did not result from excessive transfusions but may be assumed to reflect some abnormality of the iron metabolism. While one will not relinquish use of life-saving hemotherapy on account of this potential hazard, it lends re-emphasis to the need for judicious use of blood, especially in patients with hemolytic processes in whom rapid destruction of transfused blood may be anticipated.

3. **Disturbances in hemostasis (transfusional hemorrhagic diathesis).** It is well known that incompatible blood transfusions may bring about severe hemorrhagic tendencies. Since experimental studies of incompatible blood transfusions in animals demonstrated a significant drop of the fibrinogen level in addition to thrombocytopenia, administration of fibrinogen was recommended as a potentially life-saving measure in such conditions[40]. The most likely explanation for these phenomena is that intravascular coagulation occurs as a result of release of thromboplastic material from the hemolyzed red cells[41,42].

In contrast to hemorrhagic disorders of this nature, several authors called attention to the fact that massive transfusion with compatible blood may also produce severe disturbances in hemostasis, especially in patients undergoing surgery. The bleeding tendency was associated with thrombocytopenia, but could not be corrected by administration of platelet concentrates[43]. Similarly, a study of 32 patients with post-transfusional bleeding tendency disclosed them to show thrombocytopenia, especially when large quantities of blood had been administered fairly rapidly[44]. Considerable thrombocytopenia was also observed in all five infants with hemolytic disease of the newborn studied after they were subjected to replacement transfusions[43], and experimental work with exchange transfusions gave similar results[45].

Administration of fresh plasma, or blood collected in plastic containers, has been recommended for treatment of these forms of hemorrhagic diathesis[46], but it must be admitted that their etiology has not yet been fully elucidated. While some recent papers re-emphasize the significance of thrombocytopenia[47], the generalized excessive oozing of patients undergoing major surgery was described by other authors[48] to be associated with one or more of the following three factors: deficiency in prothrombin or factor V; decreased prothrombin consumption; and increased fibrinolytic activity. On the other hand, a group of investigators studying blood-clotting mechanisms in 38 patients who received massive transfusions in connection with cardiac surgery, failed to find any association between the bleeding tendency and specific disturbances of blood coagulation as shown by laboratory tests, including platelet count, fibrinogen concentration and fibrinolysin. On this basis they suggested that in such patients hemotherapy aggravated a pre-existing bleeding tendency rather than specifically interfered with one or the other hemostatic mechanism[49].

III. ISOSENSITIZATION

Since discovery of the Rh factor and general acceptance of the rule that Rh-negative recipients must not be transfused with Rh-positive blood, one of the most important causes of isosensitization by blood transfusion has been eliminated. In this connection, it is of interest to refer to the experience collected in Korean war casualties transfused exclusively with O Rh-positive blood: 50 per cent of the Rh-negative recipients were found subsequently to have become sensitized to the Rh factor[50].

While under peace-time conditions such forms of sensitization can be avoided, it would be entirely erroneous to assume that sensitization resulting from transfusions is not encountered in persons other than Rh-negative. In one series of 7,650 recipients, 107, or 1.4 per cent, showed sensitization to factors other than A and B with anti-Kell antibodies having the highest incidence[51]. Similarly, a study in our laboratory[52] demonstrated that roughly one out of five hundred Rh-positive patients proved to be sensitized to one of the Rh factors.

In this connection, special attention must be called to the consideration of O Rh-negative blood as presumably safe for transfusions without any previous cross-matching tests. It must be kept in mind that approximately 17 per cent of the U. S. Caucasoid population are susceptible to sensitization to the hr'(c) antigen present in all Rh-negative blood. Hence, administration to Rh-positive persons of such blood carries a calculated risk of either producing sensitization or even a hemolytic transfusion reaction in patients previously sensitized.

Several reports in the literature have documented dramatic examples of such events[53–55]. A near fatal hemolytic transfusion reaction has been recorded[56] which occurred after an "unnecessary" transfusion of a woman sensitized to the Duffy (Fy^a) factor.

While it is not possible to predict with any degree of accuracy occurrence of isosensitization in individual recipients, certain general statements can be made as to the likelihood for isosensitization: 1) it increases in direct proportion to the frequency of transfusions; 2) it is greater after multiple transfusions spaced in intervals of days or weeks rather than after massive transfusion therapy given within a matter of few hours or days; 3) it varies depending on the antigenicity of the blood factors; next to Rho (D), rh"(E), hr'(c), Kell (K), Duffy (Fy^a), and Kidd (Jk^a) are most commonly responsible; 4) it is influenced by the frequency of the blood factor in the population, with the greatest theoretical opportunity for isosensitization present when 50 per cent of the population is free of the blood factor and 50 per cent possess it. In this situation, a "negative" person has each time a 50 per cent chance of being transfused with "positive" blood, and since, as a rule, a minimum of two transfusions is required for sensitization, there is one chance out of four that two transfusions given in succession will be of the "positive" type for a negative recipient.

By virtue of this fact, sensitization to the Kell factor, in spite of its antigenicity, is quite rare because it is present in only 10 per cent of the population. This offers only a chance of 1 out of 100 for a Kell-negative person to receive two successive transfusions of Kell-positive blood; 5) the combination of pregnancy with transfusions increases the chances for isosensitization when the blood and fetus happen to possess the same antigen; 6) host factors are most likely involved as they affect the constitutional susceptibility to respond with formation of antibodies to antigenic stimulation.

Since obviously many of the factors enumerated are outside of our control, the basic rule is imposed on us to avoid any transfusion that is not strictly indicated in order not to expose a patient to an unnecessary risk of isosensitization.

IV. TRANSMISSION OF DISEASE

Adequate screening of blood donors prior to their acceptance is expected to rule out transmission of disease by blood transfusion. For this purpose, an adequate interview eliciting salient information about the medical history of the donor is most important. This applies particularly to prevention of post-transfusional *homologous serum jaundice* (hepatitis B). No person with any his-

tory of jaundice should be accepted as a blood donor unless the etiology of the jaundice can be unequivocally established as unrelated to hepatitis. This prohibition carries no time limitation, since carriers of hepatitis virus have been found capable of infecting recipients after indefinite periods of time[57].

Figures on incidence of post-transfusional hepatitis vary according to geographic location, number of transfusions given, and type of follow-up used. Korean casualties transfused with blood alone showed an incidence of 3.6 per cent of hepatitis[58], while reports from England indicate for recipients of blood only an incidence of approximately 0.3 per cent[59]. From Sweden an incidence of roughly 1 per cent was reported for recipients of single transfusions[60]. On the other hand, in Chile 4 per cent of 144 recipients were found to develop hepatitis, a fact possibly related to the high incidence of epidemic hepatitis in that country[61].

Undoubtedly, to have available laboratory tests capable of detecting the carrier state of hepatitis would be most desirable, but this goal remains as yet to be reached. Liver function tests, especially thymol turbidity, have been extensively investigated and found to detect a certain percentage of carriers of hepatitis virus[62,63]. Elimination from use of transfusions of donors showing elevated thymol turbidity tests appears justified on the basis of results of Jennings and associates[64] who observed a significantly higher incidence of hepatitis in recipients transfused with blood showing elevated thymol turbidity values as compared with a control series. Reduction of incidence of hepatitis after use of thymol turbidity for screening donors has also been reported by Rosenberg[65]. Strumia and associates[66] also found a significant reduction in incidence of hepatitis after such tests were introduced.

However, other authors[61] failed to notice any correlation between abnormal liver function tests, including thymol turbidity, in blood donors and their ability to transmit viral hepatitis. More recently, it has been stated that serums of patients with viral hepatitis show elevated titers of agglutinins for Rhesus monkey red cells[67,68]. However, attempts to utilize this test for screening of blood donors have not yielded promising results[69]. At present, adequate interviews are probably the most important safeguards for eliminating potential carriers of hepatitis, and as an additional precaution one may exclude donors with abnormal liver function tests, although such tests are neither specific for viral hepatitis nor capable of detecting all carriers.

Transmission of other diseases can be more readily avoided. *Syphilis* cannot be transmitted through banked blood after refrigeration of more than 72 hours. In addition, another safeguard is represented by serologic tests for syphilis, one of the routine procedures required before release of blood for transfusions. Prevention of *malarial* infection by transfusion depends mainly on attempts to ascertain from the prospective donor a history of previous malarial attack or of suppressive treatment if he has resided in a malarial area within a period of two years or less before donation. Cases of transmission of malaria have occurred even decades after a donor has left malarial areas, and even in the absence of known malarial attacks. Hence, whenever any doubt exists, blood of such persons should be diverted from transfusional use and rather utilized for preparation of plasma or other blood fractions.

A hazard of much lesser significance in our geographic area is transmission of *brucellosis* by blood transfusions. Such instances have been reported from England[70] and from South America[71]. Other diseases transmissible by transfusions include *infectious mononucleosis* according to a report from the Netherlands[72], and a viral agent which induced in recipients a brief febrile disease[73].

Considering these possibilities, it is a wise precaution to urge donors to notify the blood bank of any infectious disease that may become manifest within two weeks after blood donation, since this may make it possible to withhold transfusional use of blood from this donor. An interesting experience lending support to this recommendation was encountered in our laboratory: failure of

such notification caused transmission of viral hepatitis by blood from a donor who developed clinical symptoms of disease one week after donation[74].

V. RECIPIENTS REQUIRING SPECIFIC CONSIDERATION IN HEMOTHERAPY

While the preceding discussion dealt with hemotherapy in general, additional precautions must be taken in connection with certain patients and medical situations. One of these categories concerns patients with *auto-immune hemolytic anemia* in whom hemotherapy should be used with special discretion and sparingly[75]. Selection of blood for such patients is often difficult, and the result of transfusions disappointing unless consideration is given to the specificity of the autoantibodies which can frequently be demonstrated[76,77] and which make it necessary to select the least incompatible blood.

Another important category is *hemolytic disease of the newborn* (fetal erythroblastosis), especially in connection with replacement transfusion. It is an unfortunate misconception that O Rh-negative blood is always best, or even suitable for treatment of every instance of this disease. Actually, selection of blood is best performed with full knowledge of the identity of the antibody responsible for the disease. An indispensible confirmation of the proper selection of blood used for the infant is furnished by tests showing the blood to be compatible with the maternal serum.

Special methods for preservation and selection of blood must be used for *cardiac surgery* employing heart-lung bypass machines. Since heparinized blood has been commonly used for such purposes[78], it is obvious that special problems exist based on the need for fairly large amounts of one type of blood which can be preserved only for 24 hours. For this reason efforts have been made to find other methods for preservation of blood suitable for use in open-heart surgery, and promising approaches toward this end have been recently reported[79,80].

Since prolonged storage of blood is responsible for a significant rise in plasma potassium, it is advisable to avoid use of such blood for patients with *chronic renal disease* for whom hyperpotassemia represents a potential hazard. Recent studies also showed a rise in the ammonia content of stored blood and it was suggested that "aged" blood should not be used for patients with *cirrhosis*[81].

It must be also pointed out that patients with certain diseases are benefited much more by administration of specific blood fractions than by whole blood. This includes use of platelet concentrates for patients with severe *thrombocytopenia,* fresh frozen plasma in *hemophilia,* and especially of fibrinogen for patients suffering from severe hemorrhage caused by hypofibrinogenemia, e.g., after *abruptio placentae.*

A few comments need also be made concerning situations in which the urgency for giving blood precludes any delay, e.g., following accidents or other causes of severe hemorrhage and shock. Hence, in such conditions one must omit the usual pretransfusion tests for selection and crossmatching before the blood is administered, in order not to withhold a life-saving treatment. However, because of the calculated risk inherent in this procedure, the physician attending the patient must specifically request such *emergency transfusions,* for which so-called "safe universal donor" blood is used which is group O Rh-negative and free of irregular antibodies.

It must, however, be emphasized that the concept of group O blood being safe for recipients of all ABO groups dates back to the early days of hemotherapy and that we have since become aware of dangerous quantitative and especially qualitative properties of isoagglutinins. Although screening tests for elimination of O blood with dangerous isoagglutinins have been proposed[82] such measures are not always successful, as demonstrated by a recent experience with a hemolytic transfusion reaction resulting from use of "universal donor" blood[83]. Furthermore, there is also the possibility that antibodies for blood factors other than A and B might be present in the recipient such as anti-hr'(c)[84] or other antibodies referred to in section III, which

could be responsible for hemolytic reactions.

A final caution worthy of emphasis deals with the prohibition of specific donors for specific recipients, namely, that no woman should ever be transfused with the blood of her husband, or children. Contrary to an opinion expressed recently[85], this practice should be definitely discouraged since there is convincing evidence that exposing a woman to a "two-pronged" antigenic stimulation by transfusion with blood of her husband and by pregnancy with a fetus carrying blood factors inherited from her husband, increases the danger of isosensitization[86]. The same holds true for transfusion of a woman with blood of her children. Also, the use of the same donor for more than one transfusion carries with it a higher risk for isosensitization[87] and should be avoided.

It is a sobering thought that according to a recent estimate[88] death attributable to hemotherapy occurs in from one in 1,000 to one in 3,000 transfusions, thus exceeding the mortality after appendectomy or general anesthesia. These statistics make it necessary to combat abuses of blood transfusions in order not to jeopardize the great achievements for which hemotherapy can claim credit[89].

Probably the most important safeguard in assuring maximum benefit and minimum detriment to be derived from use of blood transfusion is the realization that hemotherapy must not be used mechanically. Specifically, the value of transfusions depends not only on fullest technical competence of blood bank personnel and adequate physical facilities but even more on the availability of a physician experienced in immunohematology and hemotherapy who can advise and assist the clinicians in all aspects of blood transfusions. Such coordinated efforts, especially if accompanied by research into as yet imperfectly understood problems of immunohematology, will not only offer protection against past and present pitfalls of hemotherapy, but also provide for future progress in this field.

REFERENCES

1. Special article: Blood transfusions in the United States. J.A.M.A., *165*:1135-1141, 1957.
2. Straus, B. and Torres, J. M.: Use and abuse of blood transfusions. J.A.M.A., *151*:699-701, 1953.
3. Davidsohn, I. and Stern, K.: Diagnosis of hemolytic transfusion reactions. Am. J. Clin. Path., *25*:381-383, 1955.
4. Wilson, H.: Acqueous dextrose solutions: a hazard in transfusions. Am. J. Clin. Path., *20*:667-668, 1950.
5. Dreyfus, B. and Salmon, C.: Inconveniences of combined infusions of citrated blood and isotonic glucose. Presse med., *60*:845-846, 1952.
6. Payne, R.: The association of febrile transfusion reactions with leuko-agglutinins. Vox Sang., *2*:233-241, 1957.
7. Brittingham, T. E.: Immunologic studies of leukocytes. Vox Sang., *2*:242-248, 1957.
8. Dausset, J., Fonseca, A. and Brecy, H.: Elimination de certain chocs transfusionnels par l'utilization de sang appauvri en leucocytes. Vox Sang., *2*:248-256, 1957.
9. Brittingham, T. E. and Chaplin, H., Jr.: Febrile transfusion reactions caused by sensitivity to donor leukocytes and platelets. J.A.M.A., *165*:819-825, 1957.
10. Marchal, G., Dausset, J., Colombani, J., Bieski-Pasquier, G. and Jaulmes, B.: Immunization anti-leukocytaire et anti-plaquettaire provoque par l'injection repetee du meme sang. Sang, *29*:549-560, 1958.
11. Ferris, H. E., Alpert, S. and Cookley, C. S.: Prevention of allergic transfusion reactions. Am. Pract., *3*:177-183, 1952.
12. Winter, C. C. and Taplin, G. V.: The value of chlor-trimeton in the management of acute allergic and febrile reactions to blood transfusion. Ann. Allergy, *12*:717-727, 1954.
13. Stephen, C. R., Martin, R. C. and Bourgeoise-Gavardin, M.: Antihistaminic drugs in treatment of nonhemolytic transfusion reactions. J.A.M.A., *158*:525-529, 1955.
14. Marchand, W. E.: Use of morphine in terminating chills and as an antipyretic. New England J. Med., *253*:315-318, 1955.
15. Davenport, J. W.: Transfusion reactions and their treatment. Am. J. Clin. Path., *24*:331-338, 1954.
16. Offenkrantz, F. M., Margolin, S. and Jackson, D.: Prevention of transfusion reactions by intravenous chlor-trimeton maleate. J.M. Soc., New Jersey, *50*:253-255, 1953.
17. Frankel, D. B.: Use of chlorprophen pyridamine maleate injection in blood transfusion. Further observations. Ann. Allergy, *13*:319-320, 1955.
18. Report of the committee of the American Academy of Allergy on the acceptability of allergic donors in the American Red Cross blood program. J. Allergy, *26*:181-186, 1955.
19. Halpern, B. N. and Besnis, M.: Action antispherocytaire de certain corps synthetiques derives de la phenothiazine. Compt. rend. Soc. Biol., *144*:759-760, 1950.

20. Greig, M. E. and Gibbons, A. J.: Possible mechanism of action by which phenothiazine derivatives preserve stored blood. Am. J. Physiol., *181*:313-318, 1955.
21. Chapin, H., Jr., Crawford, H., Cutbush, M. and Mollison, P. L.: The effects of a phenothiazine derivative (RP 3300) on red cell preservation. J. Clin. Path., *5*:91-102, 1952.
22. Jandl, J. H. and Tomlinson, A. S.: The destruction of red cells by antibodies in man. II. Pyrogenic, leukocytic and dermal responses to immune hemolysis. J. Clin. Inves., *37*:1202-1228, 1958.
23. Guynn, V. L. and Reynolds, J. T.: The use and abuse of blood transfusions. The Surgical Clinics of North America, *38*:19-30, 1958.
24. Downs, J. W.: Problem of overtransfusion in massive hemorrhage. Ann. Surg., *148*:73-80, 1958.
25. Braude, A. I., Williams, D., Siemienski, J. and Murphy, R.: Shock-like state due to transfusion of blood contaminated with gram-negative bacilli. A.M.A. Arch. Int. Med., *92*:75-84, 1953.
26. Stevens, A. R., Jr., Legg, J. S., Henry, B. S., Dille, J. M., Kirby, W. M. M. and Finch, C. A.: Fatal transfusion reactions from contamination of stored blood by cold growing bacteria. Ann. Int. Med., *39*:1228-1238, 1953.
27. Mc Entegart, M. G.: Dangerous contaminants in stored blood. Lancet, *2*:909-911, 1956.
28. Pittman, M.: A study of bacteria implicated in transfusion reactions and of bacteria isolated from blood products. J. Lab. & Clin. Med., *42*:273-288, 1953.
29. Braude, A. I., Carey, F. J. and Siemienski, J.: Studies of bacterial transfusion reactions from refrigerated blood: The properties of cold-growing bacteria. J. Clin. Investigation, *34*:311-325, 1955.
30. Braude, A. I.: Transfusion reactions from contaminated blood: Their recognition and treatment. New England J. Med., *258*:1289-1293, 1958.
31. Brucker, J. P., Stetson, J. B., Coe, R. C., Grillo, H. C. and Murphy, A. J.: Citric acid intoxication. J.A.M.A., *157*:1361-1367, 1955.
32. Howland, W. S., Bellville, J. W., Tucker, M. B., Boyar, P. and Cliffton, E. E.: Massive blood replacement. V. Failure to observe citrate intoxication. Surg., Gynec. & Obst., *105*:529-540, 1957.
33. Cookson, B. A., Costas-Durieux, J. and Bailey, C. P.: The toxic effects of citrated blood and the search for a suitable substitute for use in cardiac surgery. Ann. Surg., *139*:430-438, 1954.
34. Argent, D. E.: Citrate intoxication following rapid massive blood transfusion. Brit. J. Anesth., *29*:136-137, 1957.
35. Soulier, J. P.: Les modifications du Ca dans les transfusions massives et la circulation extracorporeale, utilisant du sang citrate. Rev. hemat., *13*:437-444, 1958.
36. Fried, W. L., Plzak, L. F., Jacobson, L. O. and Goldwasser, E.: Studies on erythropoiesis. III. Factors controlling erythropoietic production. Proc. Soc. Exper. Biol., *94*:237-241, 1957.
37. Medici, P. T., Gordon, A. S., Piliero, S., Luhby, A. L. and Yuceoglu, P.: Influence of transfusions on the erythropoietic-stimulating factor (ESF) of anemic patients. Acta haemat., *18*:325-336, 1957.
38. Cappell, D. F., Hutchison, H. E. and Jowett, M.: Transfusional siderosis: The effects of excessive iron deposits in the tissues. J. Path. & Bact., *74*:245-264, 1957.
39. Hughes, J. T. and Truelove, L. H.: Transfusional haemosiderosis simulating haemochromatosis. J. Clin. Path., *11*:128-132, 1958.
40. McKay, D. G., Hardeway, R. M., Wahle, G. M., Edelstein, R. and Tartock, D. E.: Alterations in blood coagulation mechanism after incompatible blood transfusion. Am. J. Surg., *89*:583-592, 1955.
41. Krevans, J. R., Jackson, D. P., Conley, C. L. and Hartmann, R. C.: The nature of the hemorrhagic disorder accompanying hemolytic transfusion reactions in man. Blood, *12*:834-843, 1957.
42. Pifer, P. W., Block, M. A. and Hodgkinson, C. P.: Thrombocytopenia and hemorrhage in hemolytic blood transfusion reactions. Surg., Gynec. & Obst., *103*:129-135, 1956.
43. Stefanini, M., Mednicoff, I. B., Salmon, L. and Campbell, E. W.: Thrombocytopenia of replacement transfusions; a cause of surgical bleeding. Clin. Res. Proc., *2*:61, 1954.
44. Krevans, J. R. and Jackson, D. P.: Hemorrhagic disorder following massive whole blood transfusion. J.A.M.A., *159*:171-177, 1955.
45. Mustard, J. F.: The effect of stored blood transfusions on the platelet levels in patients undergoing surgical procedures. Acta haemat., *18*:80-97, 1957.
46. Howland, W. S.: Cardiovascular and clotting disturbances during massive blood replacement. Anesthesiology, *19*:140-152, 1958.
47. Innocenzi, A.: On 3 cases of hemorrhagic diathesis after massive transfusions of whole blood. Minerva anestesiol., *23*:37-40, 1957.
48. Zucker, M. B., Siegel, M., Cliffton, E. E., Bellville, J. W., Howland, W. S. and Grossi, C. E.: Generalized excessive oozing in patients undergoing major surgery and receiving multiple blood transfusions. J. Lab. & Clin. Med., *50*:849-861, 1957.
49. Ulin, A. W., Gollub, S. W., Winchell, H. S. and Ehrlich, E. W.: Hemorrhage and massive transfusions. J.A.M.A., *168*:1971-1973, 1958.
50. Crosby, W. H.: The safety of blood transfusion in the treatment of mass casualties. Mil. Med., *117*:354-362, 1955.
51. Giblett, E. R.: A study of the incidence of blood group antibodies: a hazard in blood transfusion. (Abstract). Clin. Res. Proc., *5*:55, 1957.
52. Stern, K., Busch, S. and Buznitsky, A.: A cross-matching test using activated papain. Am. J. Clin. Path., *27*:707-713, 1957.
53. Tovey, G. H., Warren, C. P. and Wood, E. E.: "Dangerous recipients" of transfusion (incompatible reactions after transiusion of group O Rh-negative blood). Brit. M. J., *1*:813-814, 1953.
54. Grundorfer, J.: Hemolytic disease of the newborn infant caused by maternal sensitization

to the blood factor hr'(c). Blood, *8*:609-619, 1953.
55. Yahn, O., da Silva Lacaz, C. and Mellone, O.: Hemolytic accident during transfusion due to hr'(c) factor. Rev. Hosp. Clin., *10*:197-200, 1955.
56. Robinson, R., Kristinsen, J., Hunter, L., Lewis, M. and Chown, B.: The dangerous unnecessary transfusion. Canad. M.A.J., *69*:38-39, 1953.
57. Murray, R., Diefenbach, W. C. L., Ratner, F., Leone, N. C. and Oliphant, J. W.: Carriers of hepatitis virus in the blood and viral hepatitis in whole blood recipients: 2. Confirmation of carrier state by transmission experiments in volunteers. J.A.M.A., *154*:1072-1074, 1954.
58. Sborow, V. M., Giger, B. and Mann, J. D.: Incidence of hepatitis following use of pooled plasma: follow-up study of 587 Korean casualties. A.M.A. Arch. Intern. Med., *92*:678-683, 1953.
59. Wallace, J.: Homologous serum jaundice. 5th Internatl. Congress Blood Transfusions, pp. 692-696, 1955.
60. Madsen, S.: Danger of hepatitis transmission during transfusion. Ugesk. Laeger., *116*:637-641, 1954; cit. Acta Haemat., *14*:390, 1955.
61. Katz, R., Ducci, H., Bennett, H. and Rodriguez, J.: Incidence of hepatitis following transfusions of whole blood. Am. J. Clin. Path., *27*:406-421, 1957.
62. Neefe, R. J., Norris, R. F., Reinhold, J. G., Mitchell, C. B. and Howell, D. S.: Carriers of hepatitis virus in the blood and viral hepatitis in whole blood recipients. I. Studies on donors suspected as carriers of hepatitis virus and as a source of post-transfusion viral hepatitis. J.A.M.A., *154*:1066-1072, 1954.
63. Reinhold, G.: Chemical abnormalities in blood serum associated with the carrier state of viral hepatitis. Clin. Chem., *1*:3-17, 1955.
64. Jennings, E. R., Hindman, W. M., Zak, B., Reed, J. and Brines, A. O.: The thymol turbidity test in screening of blood donors. Am. J. Clin. Path., *27*:489-502, 1957.
65. Rosenberg, J.: Blood donors sreened for viral hepatitis by thymol turbidity tests. New York State J. Med., *57*:2522-2524, 1957.
66. Strumia, M. M., Burns, M. E., Sample, A. B. and McGraw, J. J.: The incidence of post-transfusion hepatitis. II. A 13-year survey including 2 years during which blood donors were screened by means of liver function studies. Am. J. Clin. Path., *30*:133-142, 1958.
67. Hoyt, R. E. and Morrison, L. M.: Reaction of viral hepatitis sera with M. rhesus erythrocytes. Proc. Soc. Exp. Biol. & Med., *93*:547-549, 1956.
68. Morrison, L. M. and Hoyt, R. E.: Hemagglutination reactions noted in viral hepatitis. J. Lab. & Clin. Med., *49*:774-778, 1957.
69. Kuhns, W. J. and Hyland, P. J.: Experiences with a hemagglutination screening test for hepatitis. Bull. Am. Assoc. Bl. Banks, *11*:366-367, 1958.
70. Wood, E. E.: Brucellosis, a hazard of blood transfusions. Brit. M. J., *1*:27-28, 1955.
71. Pinero Garcia, P. P.: Brucelosis por hemotransfusion. Dia Med., Buenos Aires, *20*:967-969, 1948.
72. De Vos, J. F. and Kiupers, F. H.: Een geval van mononucleosis infectiosa overgebracht door Bloedtransfusie. Nederl. Tijdschr. Geneesk., *95*:3036-3039, 1951.
73. Beutler, E. and Dern, R. J.: Previously unrecognized transmissible agent in human blood. Experimental and clinical studies. J.A.M.A., *159*:989-994, 1955.
74. Stern, K. and Busch, S.: An interesting experience with transmission of homologous serum jaundice. Am. J. M. Sc., *227*:559-560, 1954.
75. Davidsohn, I.: Indications and contraindications for whole blood and its various fractions. Am. J. Clin. Path., *24*:349-364, 1954.
76. Davidsohn, I. and Oyamada, A.: Specificity of auto-antibodies in hemolytic anemia. Am. J. Clin. Path., *23*:101-115, 1953.
77. Davidsohn, I. and Spurrier, W.: Immunohematologic studies in hemolytic anemia. J. A. M. A., *154*:818-821, 1954.
78. Tauxe, W. N. and Magath, T. B.: Blood banking for intracardiac surgery. J.A.M.A., *166*:2136-2139, 1958.
79. Abbott, J. P., Cooley, D. A., De Bakey, M. E. and Ragland, J. E.: Storage of blood for open heart surgery. Surgery, *44*:698-705, 1958.
80. Brown, I. W., Jr., and Smith, W. W.: Hematologic problems associated with the use of extracorporeal circulation for cardiovascular surgery. Ann. Int. Med., *49*:1035-1048, 1958.
81. Spear, P. W., Sass, M. and Cincotti, J. J.: Ammonia levels in transfused blood. J. Lab. & Clin. Med., *48*:702-707, 1956.
82. Gardner, J. M. and Tovey, G. H.: Potentially dangerous group O blood: a screening test. Lancet, *1*:1001-1004, 1954.
83. Chaplin, H., Jr.: Hemolytic transfusion reaction associated with the transfusion of "dangerous universal donor" blood. Ann. Int. Med., *43*:1334-1340, 1955.
84. Tovey, G. H., Warren, C. P. and Wood, E.E.: "Dangerous recipients" of transfusion: Incompatible reactions after transfusion of O Rh-negative blood. Brit. M. J., *1*:813, 1953.
85. Question and Answers: Husbands as blood donor for wife. J.A.M.A., *167*:1313, 1958.
86. Davidsohn, I., Stern, K., Strauser, E. R. and Spurrier, W.: Be, a new "private" blood factor. Blood, *8*:747-754, 1953.
87. Stern, K., Davidsohn, I., Jensen, F. G. and Muratore, R.: Immunologic studies on the Be^a factor. Vox. Sang., *3*:425-434, 1958.
88. Crisp, W. E.: One pint of blood. Obstet. & Gynecol., *7*:216-217, 1956.
89. Crosby, W. H.: Misuse of blood transfusions. Blood, *13*:1198-1200, 1958.

TREATMENT OF MYOCARDIAL INFARCTION *

ALDO A. LUISADA, M.D.**

Treatment of myocardial infarction is undoubtedly of interest to any physician and is still a challenge to the medical profession on account of its complexity. A listing of the most common sequence of events and of the possible complications will be given below.

The acute episode usually is initiated by the sudden, dramatic onset of severe *precordial pain*. Following this, *ventricular fibrillation* and *death* may occur. In these cases, no timely therapy is usually possible. However, ventricular fibrillation, which may be preceded by *ventricular tachycardia* and *ventricular flutter*, may occur later (though usually within the first few days) if the patient survives the initial episode. The mechanism of this paroxysm is still incompletely known. Most authors explain it with an increased excitability of the rim of normal myocardium which surrounds the ischemic area or with stimuli which arise in this area (still surviving for a certain number of hours or days) and which spread around it. However, reflex stimuli arising in the receptors of the surrounding areas, reaching the autonomic nerve centers, and affecting the intact myocardium are also probably involved and seem to cause a diffuse increase of excitability. This has been proved experimentally long ago by comparing the incidence of ventricular fibrillation following coronary ligation in animals with light anesthesia with chloralose (which has no effect on the autonomic nerve centers) with that occuring in animals with deep barbiturate anesthesia (which depresses them). More recent studies proved that thoracic sympathectomy or procaine injection of the left stellate ganglion protect the dog from ventricular fibrillation and death following coronary occlusion (Milch et al.[32]). A different theory, advocated by Beck and coworkers[1], postulates the effect of a difference in electric potential between ischemic and unaffected areas as the cause of the increase of excitability of specific areas in the rim of surrounding myocardium. Even if this is true, autonomic reflexes, causing a diffuse increase of myocardial excitability, would still be of importance.

Apart from the most fearful episode of ventricular fibrillation, other types of *arrhythmias* frequently occur, like atrial fibrillation or flutter and atrial or ventricular extrasystoles. The former may favor the formation of thrombi within the atria; the latter, if frequent, will decrease cardiac output and favor the onset of shock. A similar result may be caused by atrial flutter or fibrillation with rapid ventricular rate.

Severe tachycardia or arrhythmia, apart from other effects, decreases coronary blood flow and further impairs myocardial efficiency.

A-V block and *bundle branch block* may complicate the clinical picture of myocardial infarction. They are the result of septal involvement, in either an anteroseptal or a posterodiaphragmatic infarct.

The infarcted area, as proven by Tennant and Wiggers[2], soon becomes soft and stops contracting. Moreover, it will tend to balloon outwards at every ventricular contraction, thus absorbing part of the dynamic energy of the left ventricular contraction and reducing cardiac output. The elasticity of the infarcted area will restitute this energy in diastole, but this will only cause an increase of residual blood and of diastolic pressure within the left ventricle, with the result of an increase of pressure in the left atrium and pulmonary circuit.

The lesion of the ventricular endocardium in the area underlying a transmural infarct often causes the formation of a *mural, ventricular thrombus*. If,

* Lecture read at the V.A. Hospital, Hines, Illinois, on May 16, 1958.

** Associate Professor of Medicine and Program Director of Cardiology under a Training Grant of the National Heart Institute.

later on, parts of this become detached, *thromboembolic phenomena* will result, due to cerebral, renal, or other visceral emboli. The same complication may occur if a clot is detached from a left atrial thrombus. The importance of these thromboembolic phenomena, which occur in nearly 20% of the cases, has been recently emphasized by I. Wright and coworkers[3] and is one of the reasons for the use of anticoagulant therapy.

Episodes of *cerebral thrombosis* may occur if the patient, in addition to coronary, also has cerebral arteriosclerosis. A severe drop in arterial pressue, due to tachycardia, arrhythmia, or shock, may be the precipitating cause, so that the two episodes, cardiac and cerebral, may follow each other in rapid succession.

Rupture of the ventricular wall may occur during the early stage of an infarct causing tamponade and death. This occurs more often if the patient is not kept at rest or, worse, if he indulges in strenuous physical exertion (Jetter and White[4], Friedman and White[4]). Ventricular rupture may also occur later through gradual thinning of the scar. However, this late occurrence is exceptional.

Heart failure is frequently noted and is first revealed by pulmonary congestion and dyspnea (left heart failure), then by venous engorgement, hepatic congestion, and peripheral edema (right heart failure). The determinants of *left heart failure* are several: destruction of a certain number of left ventricular fibers; ballooning of the infarcted area (see above); impairment of the remaining ventricular wall by coronary insufficiency; and, possibly, a diffuse disturbance of the myocardium which is the result of a complex series of neurogenic, humoral, biochemical, and mechanical factors. *Right heart failure* is favored if the infarct also involves part of the right ventricular wall. It is often precipitated by the overload of the right ventricle caused by the pulmonary hypertension which follows left heart failure (Thomson and White[5]). Additional complications are the relative mitral insufficiency caused by dilatation of the left ventricle and the less common form which follows rupture of a chorda, as well as the overload of the right ventricle caused by rupture of the interventricular septum.

Paroxysmal pulmonary edema is a frequent and severe complication. While the increase of left ventricular diastolic and left atrial filling pressures are undoubtedly important, other complicating factors play an important role. One should mention the redistribution of blood with accumulation in the lungs (Sarnoff[29]) and the possible effects of certain humoral agents (histamine, serotonin) which increase capillary permeability.

Shock is another complication which requires immediate treatment and is the cause of high mortality. Shock is initiated primarily by the weakening of the left ventricle (decrease of cardiac output, forward failure). However, a real "cardiogenic shock" only occurs in cardiac arrest, ventricular fibrillation, or extreme slowing of ventricular rate (A-V block). Therefore, other elements contribute to it. Among them are reflexes arising in the ventricular wall and causing vascular collapse, alteration of the wall of the systemic capillaries similar to that described in surgical shock, and humoral elements similar to those of surgical shock (particularly ferritin, the production of which seems due to hypoxia of the liver—Chambers and Zweifach[7], Frank, Seligman, and Fine[8]). Less dramatic but still important consequences are *anxiety, nausea* or *vomiting, profuse sweating,* and a series of *metabolic disturbances* which frequently include hyperglycemia and increase of the NPN of the blood. Some of these phenomena are of reflex nature or are caused by stimuli arising in the heart and reaching the basal centers of the brain; others are probably secondary to endocrine responses and are part of an extremely severe "stress reaction" (Selye[9]).

Initial Treatment

One of the first purposes of therapy is to *alleviate pain and anxiety.* This

should be done not only for humanitarian purposes and because the patient is seeking relief but also because most drugs which relieve pain also block, to a certain extent, afferent stimuli which may be the cause of noxious reflexes. The most commonly used drug is *morphine sulfate*. This is given in dose of 10 to 15 mg (1/6-1/4 gr.) by hypodermic injection. If relief does not occur within 20 minutes, a second dose should be given. After obtaining relief, the author believes that morphine should be continued for some time in order to take advantage of its depressant effect on all nerve centers including those of the autonomic system. Following morphine administration, respiration becomes slower, drowsiness or sleep occurs, the basal metabolic rate decreases, and all voluntary or reflex actions become more sluggish. The pulse often becomes slower on account of sinus bradycardia. Thus, morphine, in addition to the relief of pain, also inhibits visceral reflexes and decreases the work of the heart. Its effect on the respiratory center (depression) and on smooth muscles (constipation, spasm of the sphincter of the bladder) should be known and watched in order to avoid unpleasant complications.

Several authors advise concomitant administration of *atropine sulfate* (0.25—0.50 mg. = 1/250—1/100 gr.) *and morphine* in order to avoid nausea or vomiting which some patients experience. The use of this drug should be encouraged, not only for this preventive reason, but also because atropine decreases the cardiac and visceral effect of severe vagal stimulation which may occur in some cases and which may contribute to a drop in blood pressure (severe bradyardia decreases cardiac output). In general, injections of morphine plus atropine at 6 hour intervals are sufficient for both analgesic and narcotic effect. However, more frequent injections may be necessary in the first 24 to 48 hours if severe pain is present.

A drug which can be used instead of morphine is *meperidine* (*Demerol*®) in doses of 50-100 mg. by subcutaneous injections. Having both morphine-like and atropine-like effects, it should *not* be associated with atropine. It may be slightly less habit-forming than morphine. On the other hand, it is far less effective than morphine in alleviating anxiety.

Another drug which may complement the effects of morphine is *phenobarbital*. It depresses the nerve centers, reduces body metabolism and cardiac work, and exerts a certain analgesic effect. On the other hand, large doses of phenobarbital cause marked peripheral vasodilatation and may favor shock. For these reasons, the following plan is suggested:

(1) Morphine plus atropine (or Demerol alone) at 3-hour intervals in the first few hours according to the severity of pain.

(2) Morphine plus atropine (or Demerol alone) at 6-hour intervals during the first 2-3 days; then at the same intervals; but in gradually reduced doses; during the 3rd to the 7th day for the purpose of decreasing cardiac work and noxious reflexes.

(3) Phenobarbital—small doses (15 mg. q.i.d.) from the 3rd to the 5th day; medium doses (30 mg. q.i.d.) from the 5th to the 10th day. Then, according to the clinical condition, smaller doses during the day, larger dose at night, until the end of the sixth week.

It is unfortunate that this scheme is not currently followed; narcotics and hypnotics are too often discontinued early, only to witness the onset of secondary complications which might have been prevented.

Rest

Rest has the purpose of decreasing the danger of rupture of the heart muscle, the incidence of heart failure, and the possible occurrence of ectopic rhythms. It is likely that the death of patients who were not put to bed was caused by either ventricular fibrillation or rupture of the heart muscle. In regard to arrhythmias and heart failure, there is no doubt that, when greater work is performed, the left ventricle fails more often and presents ectopic rhythms with greater frequency. This is why rest

should be as complete as possible and prolonged for several weeks. According to pathologists, the scar which replaces myocardial tissue in the infarcted area is usually firm by the end of the third week. Therefore, rest is most important during the first three weeks.

In the past, the patients were kept completely immobile in a supine position; they were not allowed to move their limbs or feed themselves. On the other hand, Levine and Lown[10], Mitchell et al.[11] advise to take the patient out of bed and keep him in an easy chair during the day.

Both programs are too extreme. *The immobile, supine patient* will experience congestion of the lungs and dangerous slowing of the venous circulation more frequently. This fact, due to inaction and atrophy of the skeletal muscles, may favor phlebothrombosis, a possible cause of pulmonary embolism. For this reason, even those who advocate bed rest, either ask the patient to turn from side to side several times during the day or request the nurse to change the patient's position at periodic intervals. They also invite the patient to move his legs and arms whenever he wishes. Sometimes, if the course is favorable, the patient can be propped up at meal time and, instead of being fed, can feed himself.

The *chair treatment* is often impractical for the following reasons. First, a patient in severe pain, receiving sedatives, is only semiconscious and cannot sit up without being held. A patient in shock obviously should be lying down with the head at the level of the rest of the body. On the other hand, a patient in failure with orthopnea should be propped up in bed and could be treated in a chair. However, the sitting position requires greater muscular work, a fact which increases dyspnea. Furthermore, most chair patients are difficult to convince that they must not move about and take care of their needs, especially if treated at home. On the other hand, there would be considerable theoretical advantage in using the armchair treatment for patients with mild attacks, no complications, and severe mental depression. It should be remembered that if the armchair system is adopted the patient should be lifted from bed and deposited in the chair in the morning; he should be taken from the chair and put back into bed at night. He should not be allowed to walk from the bed to the chair or jump into bed.

In regard to bodily functions, it has been proved that the *bed pan* represents a severe strain for most patients while the *commode* represents a moderate one. Again, the patient should be lifted to and from the commode; he should not be allowed to do this unassisted.

Barbiturates, in addition to decreasing the excitability of the nervous system, also decrease the wish of the patients to move about, get up, and increase their activity. This is important because many of these patients, especially in the younger age group, refuse to submit to a long period of rest and are anxious to resume normal activities. On the other hand, sedation should not be so deep as to put the patient to sleep in the daytime.

Prevention of ectopic rhythms

Another problem is represented by the increased excitability of the myocardium. *Oxygen* by inhalation has proved to be useful in this respect and is given routinely to patients with myocardial infarction. Animal experimentation (Smith and Lawson[18]) has proved that oxygen administered under pressure shortly after coronary occlusion may be lifesaving and may avert ventricular fibrillation. The mortality in 30 dogs was as follows: 20 were given air or oxygen at normal pressure (12 died; 11 with ventricular fibrillation); 10 were given oxygen at 2 atmospheres for 1 hour (only 1 died).

Oxygen is frequently given by means of an oxygen tent. In spite of the advantages of this method of administration, the above study would indicate that oxygen should be given by mask and under pressure. However, no such elevation of pressure is possible in clinical cases. Intermittent inspiratory pressure respiration obtaining a pressure of 20-30 cm. water seems advisable.

Rest and *sedatives* help in preventing

extrasystoles, paroxysmal tachycardia, atrial flutter or fibrillation, and ventricular flutter or fibrillation. Routine prophylactic use of *quinidine* has been recommended by Levine[12]. This can be questioned because quinidine is a depressant of the myocardium, lowers blood pressure, and causes sinus tachycardia. For these reasons, quinidine should be given only in cases who exhibit extrasystoles (and, therefore, have a greater likelihood of developing paroxysmal tachycardia or flutter) or present atrial flutter or fibrillation from the beginning. In these cases, usually one tablet of quinidine (gm. 0.20) q.i.d. is given, but the dose should be increased, even doubled, if severe arrhythmias develop. Larger doses should not be given if there is evidence of severe heart failure. If heart failure is present, digitalization should precede the administration of quinidine. In cases with ventricular tachycardia or ventricular flutter, one should prefer *procaine amide* (Pronestyl®) to quinidine. This is administered by the slow i.v. drip method. The effective dose is usually between 0.5 and 1 gm. The infusion should be slow because a rapid injection may cause cardiac arrest, severe vascular collapse or, paradoxically, ventricular fibrillation. It has been stated that cases of ventricular tachycardia, not previously digitalized and which do not respond to procaine amide, should be digitalized (Scherf[13]). This suggestion has been objected to with the observation that probably these cases responding to digitalis had a supraventricular tachycardia with abnormal ventricular conduction. If, on the other hand, the patient is fully digitalized and develops ventricular tachycardia, it is wise to stop digitalization, at least for a certain number of days, and try i.v. potassium with extreme care. If necessary, digitalization will be resumed after a few days.

Anticoagulants

Another problem which is still debated concerns the use of *anticoagulants*. Those who advocate the anticoagulants point out the sharp decrease of thromboembolic phenomena and the lower mortality in treated patients. Others maintain that no important difference can be found between treated and untreated patients. It seems likely that studies made in different hospitals are not exactly comparable because morbdity and mortality in a City Hospital may be higher than average. Another objection which is raised by the opponents of anticoagulant therapy is that patients treated with these drugs present a higher percentage of ventricular rupture with tamponade, or of hemopericardium due to seeping of blood from capillaries. Even if this were statistically true, the actual percentage of such cases is certainly small. Moreover, there is no question that the practical details in the application of anticoagulant therapy are very important. One cannot avoid thinking that at least some of the cases presenting hemorrhages or cardiac rupture had received too large doses of the drug.

In addition to those who advocate anticoagulants in *all* cases of infarct (and even in cases of severe coronary insufficiency) and those who advise *against* anticoagulants in general, there is a third group which accepts the use of these drugs, but only in severe cases. Anticoagulants should be used, according to them, in patients above 60, in cases with massive cardiac infarcts, shock, pulmonary edema, or peripheral congestion; and in cases with atrial fibrillation, pulmonary infarcts, or history of thrombophlebitis. Only mild cases of myocardial infarct in relatively young or middle-aged patients should not be treated with anticoagulants. Unfortunately, some cases with apparently mild episodes present thromboembolic phenomena during the course of the disease. Anticoagulant treatment might have prevented this dangerous complication.

Anticoagulant therapy was started in the early forties, largely as a result of the studies of Quick[14] and of Link[15]. It has now become an integral part of medical treatment of myocardial infarction, even though reservations and opposition are still voiced from time to time.

An important point concerns the level

at which the clotting activity should be maintained. Theoretically, a very slow clotting process is ideal, but this may cause complications. For this reason, the best level is considered to be between 25 and 30% of coagulation, as revealed by the prothrombin time and other laboratory tests.

According to most authorities[3], it is best to start treatment with a rapidly acting anticoagulant, like *heparin,* which should be given intravenously by the drip method. One dosage which can be recommended is 100 mg. of heparin in 500 cc. of 5% glucose, injected at the rate of 20 drops per minute for 12-48 hours. This slow rate is recommended in order not to overload the cardiovascular system of patients who have a potentially failing heart. Otherwise, 100 mg. of heparin can be given intramuscularly every 12 hours for 24-48 hours. Heparin prolongs clotting time but fails to modify prothrombin time. Together with heparin, a slower anticoagulant is given by mouth, so that when heparin is discontinued the oral drug will be acting. In this regard, it should be kept in mind that bishydroxycoumarin (*Dicumarol*®) initiates its effect after 48 to 72 hours; ethyl biscoumacetate (*Tromexan*®), after 18 to 24 hours, and coumadin sodium (*Warfarin*®) after about 12 hours. Depending upon which of these drugs is given, heparin is discontinued after 48, 24, or 12 hours, as soon as laboratory tests indicate that it is no longer needed.

If mild epistaxis, hematuria, or occasional petechiae appear, the dose of the anticoagulant should be decreased.

If one is using Dicumarol, it should be kept in mind that any change in dose will be reflected by clinical and laboratory changes which take place 2 or 3 days later; therefore, there is no need for frantic adjustment which only reflect the impatience of the physician.

Tromexan is definitely more difficult to handle than Dicumarol. On the contrary, Warfarin is easy to administer because its effect starts much earlier but lasts as long as that of Dicumarol.

The dose of Dicumarol is 200-400 mg. on the first day, about 200 mg. on the second, and then, once equilibrium has been reached, it may vary as much as from 25 to 100 mg. daily. With Tromexan, about four-to-five times larger doses should be given. On the first 1,000 to 1,200 mg. followed by daily doses of 200 to 600 mg. Warfarin requires a smaller dose[16,17], i.e., 50-75 mg. on the first day while the maintenance dose is 5-15 mg. daily.

One of the contraindications to the use of anticoagulants is the existence of an active gastric or duodenal ulcer, of a pulmonary cavity, or of bleeding hemorrhoids. This is why it is always necessary to inquire about a previous history of gastrointestinal disturbances and of hemorrhagic episodes. In order to obviate possible dangers, one should have available vials of *Vitamin K,* a drug which facilitates the synthesis of prothrombin. From 30 to 80 mg. of this drug are injected whenever there is evidence of severe hemorrhage. In general, the synthetic preparation menadione sodium bisulfate is used.

Treatment of Heart Failure

Another problem is represented by *heart failure.* There have been numerous discussions about treatment of this complication. There is no doubt that, apart from a propped up position and a more prolonged administration of oxygen, patients in failure may require myocardial stimulants, i.e., digitalization. Knowing that digitalis in large doses, or given too rapidly, may increase myocardial excitability and cause ectopic rhythms, a gradual digitalization should be preferred, so that complete digitalization is obtained in 36 to 48 hours. In such a way, it is possible to see whether the heart tolerates a certain dose without untoward reaction and to decrease or withdraw the drug if necessary.

Another precaution which is recommended by the author is to use a rapidly eliminated digitalis preparation, like (*Lanoxin*®), *Cedilanid dioxin or acetyldigitoxin* (*Acylanid*®). If, during the administration, one decides to discontinue the drug because of arrhythmias or paroxysmal tachycardia, digitalis will be eliminated more rapidly and the un-

desired effect will rapidly disappear.

There is no contraindication to the use of both digitalis and quinidine. One can usually give the two drugs with safety, except in cases with incomplete AV block.

Whether or not digitalis should be given when the patient is in *shock* is a problem which has not been solved. Experimental studies in animals seem to prove that digitalis decreases the severity of cardiogenic shock. On the other hand, there is experimental and clinical evidence that in the first few hours of digitalization digitalis may cause a decreased cardiac output, a fact which might increase the danger of shock.

In addition to digitalis, diuretics are frequently helpful. The most commonly used are *mercurial diuretics* or *chlorothiazide* (*Diuril*®). Their doses are similar to those used in other cases with heart failure. The hypotensive effect of chlorothiazide should be taken into consideration and the drug should not be given if blood pressure is low.

Oxygen should be administered for a longer period of time if the patient is in failure. Obviously, after the initial stage, an oxygen tent will be used. This permits an easier handling of the patient and avoids undue hardship for him.

Shock

Shock is one of the most dangerous complications. According to various statistics of a few years ago, at least 80% of patients with myocardial infarcts who went into shock died; the best that has been obtained by means of prompt and wise therapy is to cut down this percentage to about 40-50%. *It is very important to start treatment as early as possible.* An interval of six hours in the onset of therapy may mean the difference between life and death.

The danger of impending shock is largely evaluated on the basis of the blood pressure level. This, again, should be compared to the habitual level of pressure of each patient. If the patient had normal or low blood pressure, a systolic level of 80 may represent the borderline; below this, one should start preventive measures. On the other hand, if the patient had high blood pressure, a systolic pressure of 100-110 should be accepted as evidence of impending shock.

The mechanism of shock following myocardial infarction is not completely clear. As already stated, the primary factor is left ventricular failure. However, true peripheral vascular failure and hypoxia seem to be concurrent factors (Freis et al.[19]).

Therapeutic attempts, based on the premise that any increase of systemic arterial pressure would be followed by an increase of coronary flow and thus contribute to an improvement of myocardial function, include the use of vasopressor drugs (see below). However, favorable results may be due to the stimulating action of the drug on the myocardium.

The first treatment of shock which was tried was based on the use of an *intravenous infusion* of physiologic glucose, plasma, or blood. Contrasting opinions were presented (see Cochran et al.[20]—unfavorable; Gootnick and Knox[21]—favorable). Those who favor fluids advise to give them in doses of 1500 cc in 7 to 48 hours, unless there is pulmonary edema.

Intra-arterial infusions have been advocated by Silber et al.[22] and by Berman and Ackman[23]. The rationale for them was that they would permit maintenance of aortic (and coronary) pressure at the desired level with a minimum of additional volume. However, this procedure did not decrease mortality below that observed following intravenous infusions.

Among the "vasopressor drugs", one which seems to have definitely favorable results is *nor-epinephrine* (*Levophed*®). This drug increases but slightly the excitability of the myocardium, stimulates the contractility of the heart muscle, and promotes peripheral vasoconstriction. It does not cause tachycardia. The accepted method of administration is to dilute 4-8 mg. of nor-epinephrine in 500 cc of 5 per cent glucose solution, and to start a slow intravenous drip infusion. The rapidity of administration should be guided by the level of blood

pressure and by clinical signs (rales at the bases indicate increasing pulmonary edema and dictate a slower administration). If necessary, the same dose can be administered again after a few hours.

If the patient is not in shock as yet, but the blood pressure is low and tends to drop, one can use *mephentermine* (*Wyamine®*). This has a milder drug action than Levophed, has absolutely no cardiac action, and causes a moderate rise in pressure. The dose is 30 mg. in 500 cc of glucose solution, and the same dose can be repeated after four to five hours.

Another sympathomimetic amine, advocated by Sarnoff et al.[28] is *metaraminal bitartrate* (*Aramine®*). This drug has a relatively long effect after a single dose, is effective both by mouth and by injection, and does not cause arrhythmias. It causes a sustained increase of cardiac contractility and a prolonged increase of aortic pressure and coronary flow. One cc of the solution contains 10 mg. From 3 to 5 cc can be injected i.m. or i.v.

Pulmonary Edema

Treatment of this condition is more difficult when associated with myocardial infarct than in other conditions. One of the reasons is that some of the remedies which are usually employed for pulmonary edema (oxygen, digitalis, phenobarbital, morphine) are already being used. Other, frequently employed drugs and procedures (venesection, tourniquets, pressure respiration, mercurial diuretics, etc.) decrease venous return to the right heart and thus favor shock. This is one of the reasons for the study, made by the author[24], of the *antifoaming* or *defoaming therapy*. Such therapy is based on the inhalation of oxygen-ethyl alcohol vapors (95 per cent alcohol is placed in the humidifier bottle if a nasal catheter is used; 40 per cent, if a mask is used). It is a purely symptomatic therapy which, however, may serve to gain time and to interrupt a vicious circle by permitting better ventilation of the lungs.

In cases suffering from both *pulmonary edema and shock*, antifoaming therapy may enable the physician to give a larger dose of stimulating drugs, thus saving the patient by overcoming a state of reversible shock without causing an increase of foam in the respiratory passages.

Limitation of the size of the infarct; Promotion of collateral circulation

Limitation of the size of the infarct has been tried through the use of *corticosteroids*. Cortisone was found extremely effective in the experimental infarct of the dog (Johnson et al.[25]) while ACTH had no effect on size or histologic appearance of these infarcts (Wartman et al.[26]). Clinical reports are contradictory. Some useful results were obtained by Sampson and Singer[27] while none were obtained by Griffith et al.[30] and by Bergy et al.[31]. Therefore, corticosteroids are not recommended at the present time.

Papaverine hydrochloride seems to have a favorable effect in experimental infarcts by promoting collateral circulation. Clinical results have not supplied definite data. If one wishes to use this alkaloid, the dose is 100 mg. by subcutaneous injection, three or four times a day. The author considers this as a useful drug but advises its use *after* the initial period, when morphine or demerol is being discontinued, and not in the early stage. The hypotensive action of papaverine should also be considered because patients with low blood pressure may not tolerate this drug as well as the others.

Aminophyllin has been advocated following experimental studies. However, the clinical results obtained with this drug do not seem to indicate any definite action. Furthermore, oral administration is not very effective; intramuscular administration is painful; intravenous administration, even though not definitely hazardous, certainly is somewhat more risky after a recent infarct. Therefore, the author does not advocate routine use of aminophyllin.

Destruction of the blood clot

Recent studies by various groups seem to indicate that it might be possible to

Treatment of an average episode of Myocardial Infarct

	Oxygen Morphine and Atropine (or Demerol)	Phenobarbital	Papaverine	Bedrest	Anticoagulants	Complications
1st – 3rd day	Full dose	Small dose or none	----	Complete bed rest	Heparin	If heart failure – digitalis
4th – 7th day	Tapering dose	Increasing dose	----			If ectopic ventricular rhythms – quinidine or procaine amide
Second week	Small dose or none	Small, regular doses q.i.d., (plus a larger dose at night)	Small doses		Warfarin or Dicumarol	If pulmonary edema – alcohol-oxygen vapor by inhalation
Third week	----		Larger doses	Sitting in bed		
Fourth week	----	At night only		Initial mobilization		If shock – nor-epinephrine (Levophed) (Wyamine or Aramine)
Fifth week	----			Easy chair. A few steps		
Sixth week	----			Convalescence		

dissolve the clot of a coronary thrombus by using either plasmin (an enzyme which is normally found in the blood) or *stroptokinase* (an enzyme produced by the streptococcus hemolyticus which stimulates production of plasmin within the body). It is too early for an evaluation of these methods. Should their effectiveness and lack of side effects be confirmed, then the early treatment should consist of, first, destruction of the clot through enzymes, and then prevention of a new clot by anticoagulants.

Diet

The diet will be based on fluids in the first 24-48 hours. It should be *bland* for the first week. It should be a *salt poor diet* for the first three weeks, even if no evidence of heart failure is present. After the third week, general clinical considerations dictate the type of diet.

General Scheme of Therapy

A scheme of treatment of an average case of myocardial infarct is given below. It should be well understood that modifications of the scheme are necessary according to the severity of the clinical picture and according to the various possible complications.

BIBLIOGRAPHY

1. Beck, C. S. and Leighninger, D. S.: J.A.M.A., *159*:1264, 1955.
2. Tennant, R. and Wiggers, C. J.: Am. J. Phys., *112*:351, 1935.
3. Wright, I., Marple, C. A. and Beck, D. F.: *Myocardial Infarction*. New York, Grune and Stratton, 1954.
4. Jetter and White; Friedman and White: Quoted in P. D. White, *Heart Disease*. New York, Macmillan, 1951.
5. Thompson, W. P. and White, P. D.: Amer. Heart Jr., *12*:641, 1936.
6. Luisada, A. A.: Circulation, *13*:113, 1956.
7. Chambers, R. et al.: Am. J. Phys., *139*:123, 1943. Also: Zweifach et al., Ann. Surg., *120*: 232, 1944.
8. Frank, H. A., Seligman, A. M. and Fine, J.: J. Clin. Invest., *25*:22, 1946.
9. Selye, H.: *Stress: The Physiology and Pathology of Exposure to Systemic Stress*. Montreal, Acta Inc., 1950.
10. Levine, S. A. and Lown, B.: J.A.M.A., *148*: 1365, 1952.
11. Mitchell, A. M., Dealy, J. B., Lown, B. and Levine, S. A.: J.A.M.A., *155*:810, 1951.
12. Levine, S. A.: *Clinical Heart Disease*. Philadelphia, Saunders, 1958.
13. Scherf, D. and Boyd, L. J.: *Clinical Electrocardiography*. Philadelphia, Lippincott, 1946.
14. Quick, A. J.: Am. J. Phys., *118*:260, 1937; and Penna. M. J., *43*:125, 1939.
15. Link, K. P.: Harvey Lect. Series, *34*:162, 1943, and Fed. Proc., *4*:176, 1945.
16. Pollock, B. E.: J.A.M.A., *161*:404, 1957.
17. Porter, R. R., Richardson, D. and Mauck, H. P.: Virg. M. Monthly, *85*:465, 1958.
18. Smith, G. and Lawson, D. A.: Scott. M. J., *3*:346, 1958.
19. Freis, E. D. et al.: J. Clin. Invest., *31*:131, 1952.
20. Cochran, B., Wallace, W. B. and Griffith, G. C.: Amer. Therap. Soc., June, 1952.
21. Gootnick, A. and Knox, F. H., Jr.: Circul., 7:511, 1953.
22. Silber, E. N. et al., J.A.M.A., *147*:1626, 1951.
23. Berman, E. F. and Ackman, L. C.: Am. Heart J., *43*:264, 1952.
24. Luisada, A. A.: Circul., *2*:872, 1950.
25. Johnson, A. S. et al.: Circul., *7*:224, 1953.
26. Wartman, W. B. et al.: Circul. Res., *3*:496, 1955.
27. Sampson, J. J. and Singer, I. M.: Am. Heart J., *38*:54, 1949.
28. Sarnoff, S. J. et al.: Circul., *10*:84, 1954.
29. Sarnoff, S. J. and Sarnoff, L. C.: Dis. Chest, *22*:685, 1952.
30. Griffith, G. C. et al.: Circul., *9*:527, 1954.
31. Bergy, G. C., Burroughs, B. W. and Bruce, R. A.: Am. J. Med. Sci., *232*:513, 1956.
32. Milch, E. et al. Am. Heart J., *50*:483, 1955.

PRINCIPLES OF QUANTITATIVE AUTORADIOGRAPHY*

Institute of General Pathology, University of Milano, Italy

G. GUIDOTTI, M.D.**

Photographic emulsions for measuring radioactivity have not been used as commonly as other recording devices, such as Geiger or scintillation counters, partly because the emulsions are difficult to handle, and partly because the methods are difficult to standardize and reproducible quantitative data difficult to obtain. However, many of these difficulties have recently been overcome and another suitable instrument has been added to the more commonly employed recording devices.

A brief summary of photographic emulsion characteristics and of autoradiographic techniques, a review of the methods suited for quantitative assessment of autoradiographs and an outline of the effectiveness and limitations of the autoradiographic method in performing relative and absolute measurements of radioactivity will be the subject of the following pages.

The Photographic Emulsion

Nature of the Emulsion. The photographic emulsion is a dispersion of silver halide crystals in a gelatin gel. Silver halide crystals (AgBr sometimes mixed wtih small amounts of AgI) represent about 30-60% of common photographic emulsions by weight. In nuclear emulsions this percentage is increased to 80% or more. Crystal dimensions range between 0.1 μm in the finest nuclear emulsions and a few μm in some x-ray films.

The Photographic Process. Each crystal bromide acts as a unit with respect to the photographic process. When the crystal is exposed to light photons or to other ionizing radiations, a certain number of electrons is liberated within the crystal lattice. Since crystals always contain imperfections called "sensitivity specks" which behave as "electron traps," a negative electrostatic atmosphere is produced in some regions of the crystal. Positively charged silver ions, produced by the thermal agitation and present in the crystal lattice, migrate to the sensitivity specks and neutralize their negative charge with precipitation of metallic silver. This metallic silver forms the so-called "latent image."

The latent image is too small to be seen with a visible light microscope, but it serves as a nucleus for the subsequent reduction of the entire crystal by the developing solution. At the end of the process developed crystals, usually called "grains," can reach a size greater than the undeveloped ones. The chemical reaction of reduction takes place and proceeds on the surface of all crystals, but its rate is sufficiently rapid to lead to the formation of a grain only if a latent image has been produced. This is due to the catalytic action of the latent image which lowers the activation energy required for the reduction.

After development, the crystals without the latent image must be removed from the emulsion to make the image visible and to preserve it. This process is called fixation and it is carried out by sodium thiosulphate (hypo) which dissolves the crystals of silver bromide leaving the metallic silver to mark the path of the ionizing radiation.

Sensitivity of the Emulsion. The sensitivity of a photographic emulsion is determined by that of the single silver halide crystal (the inverse of the quantity of energy required to form the latent image) and by the size and concentration of the crystals. Sensitivity is an important characteristic of each emulsion as it determines the possibility of recording different ionizing particles: α-particles are recorded by a large number of photographic emulsions because their high

* Received December 1, 1958.

** Research Fellow in Physiology, The Chicago Medical School, Chicago, Illinois.

ionizing power*, while $\beta-$ and $\beta+$ ionize much less per unit path and are completely recorded only by very sensitive emulsions. In this respect, all the recently produced nuclear emulsions (Ilford G5, K5 and L4; Kodak Ldt. NT.4; Eastman Kodak NTB-3) are sensitive to "minimum ionization"** and thus record the complete track of any charged particle.

Speed of the Emulsion. The "speed" or relative sensitivity of an emulsion can be defined as the ratio between an arbitrary constant, e.g., a standard optical density above fog***, and the exposure time required to achieve it. This characteristic is useful in comparing the suitability of different emulsions for a given purpose.

Other factors affecting sensitivity are temperature, humidity, pressure and fading of the latent image.

Elements of Autoradiographic Technique

Autoradiography is a method of recording ionizing particles and can be classified under the terms of *contrast autoradiography and track autoradiography.* The former takes advantage of the differences in blackening(grain density) between different areas of the specimen; the latter records the single tracks produced by the ionizing particles emerging from the source as they hit succeeding halide crystals.

*The ionizing power of α particle depends on the square of its charge, is a decreasing function of its velocity (except for very high velocities) and is independent on its mass. It can be defined as the number of ion pairs produced per unit length along its path by the inelastic collisions with atoms of the medium through which it passes. For α given medium the ionizing power is directly proportional to the energy loss of the particle due to these collisions, per unit of path length.

** When the velocity of α particle reaches 90% of that of light, the ionizing power depends only upon the square of its charge; α single charged particle with this velocity has the lowest ionizing power that α charged particle can possess (minimum ionization). An emulsion giving latent image with such α particle, is therefore sensitive to minimum ionization.

***The optical density depends not only upon the number, but also upon the size of the developed grain; fog is represented by α halo of grains which are developed independently of the ionizing action.

Resolving Power. In general, the resolving power might be defined as the smallest distance between two point sources within a given specimen at which two distinct images can be obtained. This depends upon numerous factors, such as 1) The spacial relationship between the radioactive source and emulsion (size and thickness of the source, thickness of the sensitive emulsion layer, gap between the two); 2) Energy and intensity of the radiation (quality and quantity of the ionizing particles); 3) Back-scattering from the material surrounding the sensitive layers; 4) Characteristics of the photographic emulsion (grain size and homogeneity, silver concentration); and 5) Method of processing of the emulsion.

The best methods attain a resolving power as high as 1-2 μm or better.

For greater details the reader is referered to other papers[15, 23, 42, 48, 50, 58, 72, 82, 89].

Procedures.

A. **Apposition Method**[6, 55] (Fig. 1). In this procedure, source and photographic emulsion are kept in close contact by pressure. A thin protective layer may sometimes be interposed between source and emulsion[29]. After a convenient exposure time, the film is removed and processed. The method is used primarily with gross specimens (cut organs, smooth hard materials, paper electrophoresis and chromatographic strips,

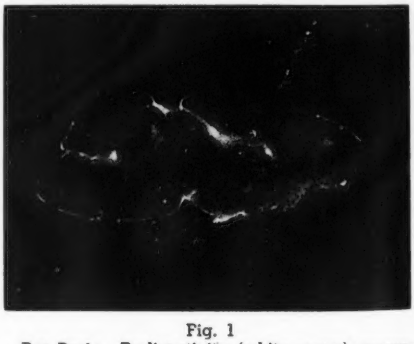

Fig. 1

Rat Brain. Radioactivity (white areas) concentrated at the periphery and along the fissures of the brain (5 X). Apposition method autoradiography. Radioactive source C^{14} ($\beta-$) injected as leucine-1-C^{14}. Emulsion: Kodirex. Exposure: 28 days.

etc.) when an approximate localization of radioactive areas is required.

B. **Mounting Method**[27,28] (Fig. 2). In this method, the specimen (histologic sections, blood, etc.) is mounted directly on the surface of the photographic emulsion to which it remains bonded throughout photographic processing and specimen staining.

A combination of apposition and mounting methods (sandwich method) is sometimes employed[36].

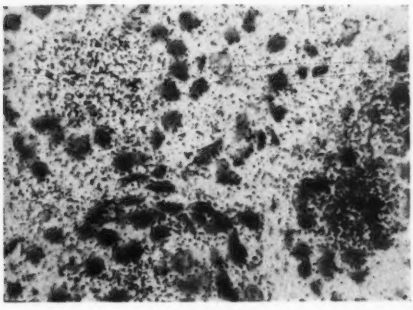

Fig. 2

Rat Thyroid. Radioactivity (black grains) concentrated in colloid (570 X). Mounting method autoradiography. Radioactive source: I^{131} (β- and γ) injected as sodium iodide. Emulsion: EK medium lantern slide. Exposure: 6 days. Staining: Hematoxylin-Eosin.

C. **Coating Methods.**

1. **Stripped emulsions.** (Fig. 3). In this method the specimen (histologic sections, smears, etc.) is covered either with a thin (stripping film autoradiography) [23,46,50,66,79,86,87] or with a thick[45] photographic emulsion which is stripped from its support.

2. **Liquid emulsions.** (Fig. 4-5). In this method the specimen is covered either with a thin[8,9,10,26,42,59,69] or with a thick[33,39,43,44,48,54,61] layer of emulsion melted from plates, films or gels.

In special cases, small objects such as cells[62], microorganisms[21,48], virus and macromolecules[64,65], inorganic crystals[63], wires[70], fine capillary tubes[11], and rocks[83] are directly mixed into the liquified emulsion. Coating methods give the best resolving powers obtainable in autoradiography.

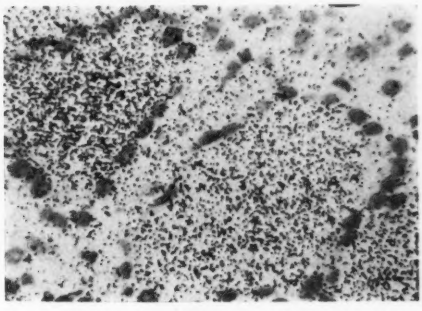

Fig. 4

Rat Thyroid. Radioactivity concentrated in colloid (570 X). Liquid emulsion (thin layer) autoradiography. Radioactive source: I^{131} (β-, γ) injected as sodium iodide. Emulsion: Ilford G5 (5 μm thickness). Exposure: 6 days. Staining: Carmalun.

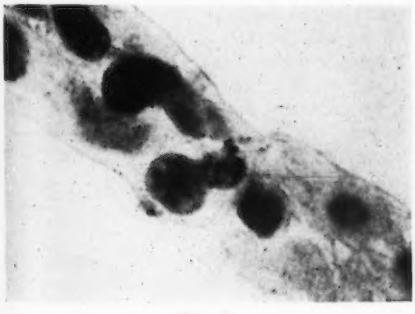

Fig. 3

Rat Lung. Radioactivity detectable over one cell nucleus (2300 X). Coating method. Stripping film autoradiography. Radioactive source: H^3 (β-) injected as labelled thymidine. Emulsion: Kodak AR 10 (4 μm thickness). Exposure: 3 days. Staining: Hematoxylin-Eosin.

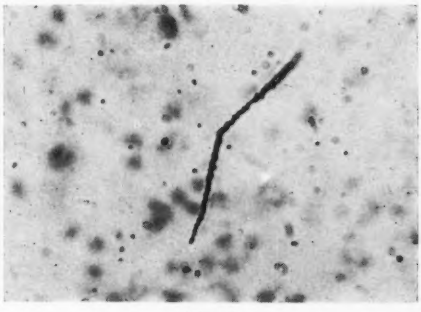

Fig. 5

Rat Spleen. Two α-tracks emerge from the tissue (570 X). Liquid emulsion (thick layer) autoradiography. Radioactive source: Th^{228} (α, γ) injected as thorium dioxide. Emulsion: Ilford C2 (50 μm thickness). Exposure: 40 hours. Staining: Pappenheim-Unna.

Autoradiography as a Method for Measuring Radioactivity—Quantitative Autoradiography

1. **General.** Radioactivity measurements by means of autoradiography are based on the estimation of the effects of ionizing radiations on the photographic emulsion.

For relative measurements, the simple record of these effects may be sufficient. For absolute measurements of the number of radioactive atoms present in a given source one must know the ratio between the number of effects (developed grains or tracks) in a given emulsion and the effective number of events (disintegrations) at the source.

2. **Collection of Raw Data.** The interaction between ionizing particles and photographic emulsion leads to a deposition of silver particles either in a diffuse form or by tracks, depending upon the emulsion characteristics. In the first case (contrast autoradiography) the phenomenon can be measured by densitometry or by counting the number of grains; in the second case (track autoradiography), by counting the number of tracks.

a. **Densitometry.** Measurements of high photographic density can be made with great accuracy by means of photometers and microphotometers (densitometers)[12, 17, 76]. The enlarged image of the required area is projected and its density measured with a photocell or phototube. Apposition autoradiography (see A. above) allows the best densitometric measurements[77].

If the density of the grains is low or a quantitative record of their distribution in selected areas of a photographic emulsion is desired, one must resort to the use of special methods such as the methods described by Mazia et al[68], and by Gullberg[49]. Since the underlying specimen does not influence the grain image, these microdensitometric methods are suitable for the utilization of high resolution autoradiographic techniques.

b. **Visual Counting of Grains.** Visual grain counting is commonly carried out at high magnification by means of a conventional microscope fitted with a vertical and two horizontal orthogonal movements controlled by micrometric screws. A suitable micrometer placed in the ocular outlines the area in which grains are to be counted. The counting can be performed either through the entire depth of the emulsion[1, 20, 52, 56, 57, 75, 78, 81] or at distinct levels in the emulsion over the source surface[37, 73, 74]. In the first case the emulsion thickness must be strictly uniform, in the second one this condition is unnecessary. Owing to these considerations, all high resolution autoradiographic methods employing thin emulsion layers are suitable for these measurements. An objection to grain counting is the tediousness of the method which may lead to a decrease in efficiency of the operator after a few hours of work; however, with proper care, the error due to personal inaccuracy does not exceed ± 10%[73], which is of the same order of that reported[68] for a microdensitometric procedure.

c. **Automaitc Grain Counting.** To avoid the very laborious task of visual grain counting, an electronic grain counter based on the "flying spot" principle has been designed by Dudley and Pelc[25]. The counter, suited for stripping film autoradiography, offers definite advantages as high sensitivity, speed and freedom from subjective errors. However, the complexity of the instrument does not allow routine use as yet[80].

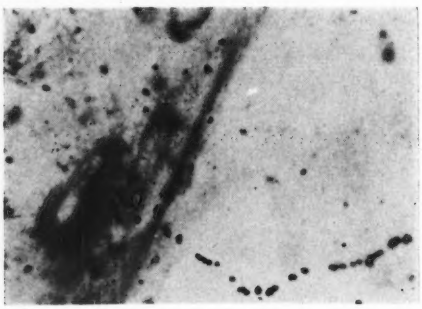

Fig. 6

Tracheal Cartilage of Rat. An electron track originates from the cytoplasm of a cartilage cell near the perichondrium (1400 X). Liquid emulsion (thick layer) autoradiography. Radioactive source: S^{35} (β^-) injected as sodium sulphate. Emulsion: Ilford G5 (100μm thickness). Exposure: 24 hours. Staining: Alcian blue 8GN.

d. **Track Counting.** Track counting in thick photographic emulsions is performed at high magnification by means of special microscopes fitted with ocular micrometers. In the photographic emulsions, α-particles produce heavy straight (or slightly distorted) tracks (Fig. 5), which can be easily recognized and counted under the microscope[27,51,60,61,71,85]. β-particles (Fig. 6), because of their small mass, are readily scattered by the atoms along their path and often give rise to irregular curly tracks[14,18,38,47,54,62]. Nevertheless, these can be counted with acceptable accuracy if their concentration per unit area is not too great.

A definite advantage of the track method lies in the possibility of identifying and excluding from the count almost all spurious tracks which either do not emerge from the specimen or possess an energy higher than the maximum of the involved radioelement[31]. Furthermore, high resolving powers obtained by nuclear track procedure make them particularly useful in the investigation of small bodies and complex structures[48].

3. **Relative Measurements.** Raw data, collected by one of the methods described above, are sufficient as simple activity indices of different parts of the source[3,4,13,16,30,32,34,40,41,47,75,78]. However, the following corrections must be introduced.

a. **Sensitometric Correction.** Since photographic density (or number of developed grains) and amount of radiation exposure (integrated product of radiation intensity by duration of exposure), are linearly related only for a certain interval, the raw data must be corrected using a calibration curve obtained with standard sources of the radiation under study[5,7,13,17,22,24,35,37,53,88,90].

b. **Background.** This correction is obvious but not always easy. Spurious events occurring within the emulsion (due to natural radiation or "cross-fire" from the utilized radiation[58]) must be evaluated and subtracted from the values after sensitometric correction.

4. **Absolute Measurements.** Absolute measurements in terms of disintegrations per unit time or of atoms present in a given volume of source at a given time are obviously more complex. Additional factors concerning emulsion, tracer and source must be taken into account; but these are not always easy to evaluate. Among these are:

a. **Fading of the latent image.** It is known that several photographic emulsions (especially the fine-grained nuclear emulsions) are liable to show appreciable fading of the latent image when time exposure is sufficiently long. This introduces serious errors in quantitative work. When complete prevention of this phenomenon[2,65,84] is impossible, a correction factor can be calculated.

b. **Self-absorption and scattering.** Self-absorption and scattering of the particles within the source and in the gap between source and emulsion usually are disregarded in the calibration curves of sensitometric correction. On the other hand, owing to the thickness of the source and of the gap, these phenomena must be taken into account when absolute measurements are desired[67,85,92].

c. **Back-scattering.** With particles of relatively high energy (or with very thin source layers), one must also take into account the effect of the particles which are emitted from the radioactive source in the direction of the supporting material and are scattered back through the source into the photographic emulsion[7].

d. **Geometrical and physical characteristics of the source.** The previously described corrections presuppose a strict homogeneity of the geometrical (shape and dimensions) and physical (average density) characteristics of all parts of the source. If this is not true, events occurring in different regions of the source may produce different effects because of variation in self-absorption and scattering of the ionizing particles. In biological materials, strict homogeneity of the characteristics of the source is difficult to obtain. Therefore, deviations from the ideal condition have to be evaluated in each case. Experiments on this problem, by means of suitable models, have recently been started[19,58,63,91].

e. **Number of radioactive atoms.** Whatever the method used for computing disintegrations, simple calculations permit the evaluation of the number of radioactive atoms present in a given volume of the source at a given time, since the number of disintegrations is proportional to the number of atoms of the radioelement present.

5. **Conclusion.** In conclusion, it can be pointed out that, in spite of limitations and difficulties, the autoradiographic technique, if correctly employed, represents a useful instrument for the measurement of radioactivity. The importance of this instrument is even higher in view of the fact that one or a few radioactive disintegrations within a source of small dimension is sufficient to produce measureable effects on the photographic emulsion and this permits a quantitative assay many orders of magnitude more sensitive than the most sensitive microanalytical techniques.

SUMMARY

Characteristics of the photographic emulsions and elements of autoradiographic technique have been summarized. Applicability of densitometry, visual and automatic grain- and track-counting to quantitative assessment of autoradiographs has been described. Finally, effectiveness and limitations of the autoradiographic method in performing relative and absolute measurements of radioactive sources have been outlined.

REFERENCES

1. Abercrombie, M. and Causey, G.: Nature. *166*:229, 1950.
2. Albouy, G. and Faraggi, H.: J. Phys. Radium. *10*:105, 1949.
3. Andresen, N., Chapman-Andresen, C. and Holter, H.: C. R. Trav. Lab. Carlsberg, Sér. Chim. *28*:189, 1952.
4. Andresen, N., Chapman-Andresen, C., Holter, H. and Robinson, C. V.: C. R. Trav. Lab. Carlsberg, Sér. Chim. *28*:499, 1953.
5. Arnold, J. S., Johnson, K. E. and Jee, W. S. S.: Feder. Proc. *13*:421, 1954.
6. Axelrod, D. J. and Hamilton, J. G.: Amer. J. Pathol. *23*:389, 1947.
7. Beischer, D. E.: Nucleonics. *11*: (n°12) 24, 1953.
8. Bélanger, L. F. and Leblond, C. P.: Endocrinology. *39*:8, 1946.
9. Bélanger, L. F.: Anat. Rec. *107*:149, 1950.
10. Bélanger, L. F.: Nature. *170*:625, 1952.
11. Bonetti, A. and Occhialini, G. P. S.: Nuovo Cimento. 8-Serie *9*:725, 1951.
12. Boström, H., Odeblad, E. and Friberg, U.: Arch. Biochem. Biophys. *38*:283, 1952.
13. Boström, H., Odeblad, E. and Friberg, U.: Acta Pathol. Microbiol. Scand. *32*:516, 1953.
14. Boyd, G. A. and Levi, H.: Science. *111*:58, 1950.
15. Boyd, G. A.: "Autoradiography in Biology and Medicine," Academic Press, Inc., New York, 1955.
16. Brachet, J. and Ficq, A.: Arch. Biol. *67*:431, 1956.
17. Branson, H. and Hansborough, L. A.: Science. *108*:327, 1948.
18. Campbell, D.: Nature. *167*:274, 1951.
19. Cormack, D. V.: Brit. J. Radiol. *28*:450, 1955.
20. Cronkite, E. P., Fliedner, T. M., Bond, V. P., Rubini, J. R., Brechner, G. and Quastler, H.: A/Conf. 15/P/840. Genève, 1958.
21. Demers, P. and Fredette, V.: Phys. Rev. *72*: 538, 1947.
22. Dockum, N. L. and Healy, J. W.: Stain Technol. *32*:209, 1957.
23. Doniach, I. and Pelc, S. R.: Brit. J. Radiol. *23*:184, 1950.
24. Dudley, R. A. and Dobyns, B. M.: Science. *109*:327, 1949.
25. Dudley, R. A. and Pelc, S. R.: Nature. *172*: 992, 1953.
26. Edwards, L. C.: Rev. Sci. Instr. *26*:515, 1955.
27. Endicott, K. and Yagoda, H.: Proc. Soc. Exptl. Biol. Med. *64*:170, 1947.
28. Evans, T. C.: Proc. Soc. Exptl. Biol. Med. *64*:313, 1947.
29. Everett, N. B. and Simmons, B. S.: Anat. Rec. *117*:25, 1953.
30. Ficq, A.: Experientia. *9*:377, 1953.
31. Ficq, A.: J. Embryol. Exptl. Morphol. *2*:204, 1954.
32. Ficq, A., Gavosto, F. and Errera, M.: Exptl. Cell Research. *6*:69, 1954.
33. Ficq, A.: Arch. Biol. *66*:509, 1955.
34. Ficq, A. and Brachet, J.: Exptl. Cell Research. *11*:135, 1956.
35. Fitzgerald, P. J., Eidinoff, M. L., Knall, J. E. and Simmel, E. B.: Science. *114*:494, 1951.
36. Fitzgerald, P. J., Simmel, E. B., Weinstein, J. and Martin, C.: Lab. Invest. *2*:181, 1953.
37. Forro, F. Jr.: Exptl. Cell Research. *12*:363, 1957.
38. Gavosto, F. and Ficq, A.: Nature. *172*:406, 1953.
39. Gavosto, F. and Ficq, A.: Ann. Inst. Pasteur. *86*:320, 1954.
40. Gavosto, F., Ficq, A. and Errera, M.: Exptl. Cell Research. *6*:238, 1954.
41. Gavosto, F. and Rechenman, R.: Biochim. Biophys. Acta. *13*:583, 1954.
42. Gross, J., Bogoroch, R., Nadler, N. J. and Leblond, C. P.: Am. J. Roent. Rad. Ther. *65*:420, 1951.
43. Guidotti, G.: Riv. Istoch. nor. pat. *1*:471, 1955.
44. Guidotti, G.: Exptl. Cell Research. *10*:544, 1956.
45. Guidotti, G. and Levi Setti, R.: Stain Technol. *31*:57, 1956.

46. Guidotti, G. and Passalacqua, F.: Experientia. *12*:117, 1956.
47. Guidotti, G.: Exptl. Cell Research. *12*:659, 1957.
48. Guidotti, G.: 2e. Colloque de Photograhpie Corpuscolaire. Montreal, 1958. Communication BY.
49. Gullberg, J. E.: Exptl. Cell Research. Suppl. *4*:222, 1957.
50. Herz, R. H.: Nucleonics. *9*: (n°3) 24, 1951.
51. Hoecker, F. E. and Roofe, P. G.: Radiology. *56*:89, 1951.
52. Hornsey, S. and Howard, A.: Ann. N. Y. Acad. Sci. *63*:915, 1956.
53. Hurwitz, J. K.: Rev. Sci. Instr. *26*:518, 1955.
54. King, D. T., Harris, J. E. and Tkaczyk, S.: Nature. *167*:273, 1951.
55. Lacassagne, A. and Lattes, C.: J. Radiol. Electrol. *9*:1, 1925.
56. Lajtha, L. G.: Nature. *180*:1048, 1957.
57. Lajtha, L. G. and Kumatory, T.: Nature. *180*: 991, 1957.
58. Lamerton, L. F. and Harriss, E. B.: J. Photogr. Sci. *2*:135, 1954.
59. Leblond, C. P., Percival, W. L. and Gross, J.: Proc. Soc. Exptl. Biol. Med. *67*:74, 1948.
60. Levi, H.: Biochim. Biophys. Acta. *7*:198, 1951.
61. Levi, H.: Nature. *171*:123, 1953.
62. Levi, H.: Exptl. Cell Research. *7*:44, 1954.
63. Levi, H.: Exptl. Cell Research. Suppl. *4*:207, 1957.
64. Levinthal, C.: Proc. Natl. Acad. Sci. *42*:394, 1956.
65. Levinthal, C. and Thomas, C. A. Jr.: Biochim. Biophys. Acta. *23*:453, 1957.
66. MacDonald, A. M., Cobb, J. and Solomon, A. K.: Science. *107*:550, 1948.
67. Mayr, G.: Tumori. *42*:411, 1956.
68. Mazia, D., Plaut, W. S. and Ellis, G. W.: Exptl. Cell Research. *9*:305, 1955.
69. Messier, B. and Leblond, C. P.: Proc. Soc. Exptl. Biol. Med. *96*:7, 1957.
70. Meulemans, G., Occhialini, G. P. S. and Vincent, A. M.: Nuovo Cimento. 8-Serie *9*:341, 1951.
71. Miller, B. L. and Hoecker, F. E.: Nucleonics. *8*: (n°5) 44, 1951.
72. Nadler, N. J.: Canad. J. Med. Sci. *29*:182, 1951.
73. Nadler, N. J.: Am. J. Roent. Rad. Ther. *70*: 814, 1953.
74. Nadler, N. J., Leblond, C. P. and Bogoroch, R.: Endocrinology. *54*:154, 1954.
75. Odeblad, E.: Exptl. Cell Research. *2*:574, 1951.
76. Odeblad, E.: Acta Radiol. Suppl. *93*:1, 1952.
77. Odeblad, E.: Acta Radiol. *45*:323, 1956.
78. Pantelouris, E. M.: Exptl. Cell Research. *14*: 584, 1958.
79. Pelc, S. R.: Nature. *160*:749, 1947.
80. Pelc, S. R.: Exptl. Cell Research. Suppl. *4*: 231, 1957.
81. Pelc, S. R.: Exptl. Cell Research. *14*:301, 1958.
82. Pelc, S. R.: Nuclear Handbook, Section: Autoradiography. G. Newnes Ltd., London, 1958.
83. Picciotto, E. E.: Bull. Centre Phys. Nucl. Univ. Libre (Bruxelles) n°33, 1952.
84. Ray, R. C. and Stevens, G. W. W.: Brit. J. Rad. *26*:362, 1953.
85. Rotblat, J. and Ward, G. B.: Nature. *172*:769, 1953.
86. Schlesinger, M. J., Levi, H. and Weyant, R.: Rev. Sci. Instr. *27*:969, 1956.
87. Simmel, E. B., Fitzgerald, P. J. and Godwin, J. T.: Stain Technol. *26*:25, 1951.
88. Sinclair, W. K., Abbatt, J. D., Farran, H. E. A., Harriss, E. B. and Lamerton, L. F.: Brit. J. Radiol. *29*:337, 1956.
89. Stevens, G. W. W.: Brit. J Radiol. *23*:723, 1950.
90. Taylor, J. H., McMaster, R. H. and Caluya, M. F.: Exptl. Cell Research. *9*:460, 1955.
91. Verly, W. G., Firket, H. and Hunebelle, G.: A/Conf. 15/P/323. Genève, 1958.
92. Yagoda, H.: "Radioactive Measurements with Nuclear Emulsions." J. Wiley & Sons, Inc., New York, 1949.

TREATMENT OF CONJUNCTIVITIS

MARTHA RUBIN FOLK, M.D.*

Conjunctivitis or "pink eye" is a relatively common condition which every physician is frequently called upon to treat. It is the purpose of this article to call to your attention certain principles of treatment. Most cases of purulent conjunctivitis are caused by the streptococcus, staphylococcus, pneumococcus or Koch's Weeks bacillus. These cases are basically self limited diseases and are frequently associated with upper respiratory infections. Since the condition is self limited, it is obviously desirable that we do not use agents which can make the condition worse. It is important to take cultures and smears for proper treatment in purulent types of conjunctivitis. One can treat such cases better by knowing the type of bacteria.

A patient with a purulent conjunctivitis has a chief complaint of having a red eye with profuse discharge. As a result of this discharge, the patient has the classical symptom of agglutination of the lids most marked in the morning. The patient may also complain of a granular feeling in his eyes. There is no pain or sensitivity associated with the above condition.

Pain, sensitivity to light, and visual loss are not symptoms of conjunctivitis. In the presence of any of these three symptoms I would suggest that the patient be referred to an ophthalmologist. These symptoms suggest the possibility of an onset of acute glaucoma, acute iritis or a virus type of conjunctivitis. These conditions are best treated by an ophthalmologist.

Many antibacterial drugs are vailable for treatment of conjunctivitis and it is important that the physician use certain principles in the selection of the proper drug.

First let us consider methods of administration. Most drugs can be administered orally, parenterally or topically. Many studies have demonstrated that the topical administration results in the highest concentration of effective agents at the site of infection in conpunctivitis. As a result, there is no need for systemic medication in the surface disease.

Topical preparations are available in drops and ointment form. Drops are more frequently used and do not have the disadvantage of causing a film over the cornea which tends to blur the vision. Ointment causes blurring but maintains a higher concentration of antibacterial agent in the conjunctival sac for longer periods of time. A common practice is to order drops during the day at hourly intervals and ointment at night. Ointments alone may be used every three hours.

Since topical medication is all that is required, we prefer to use drugs that are unlikely to be used systemically. Should the patient become sensitive to the drug, we have not deprived the patent of a drug which may be later used for a systemic illness. Recall, too, that conjunctivitis is a self limited disease and it would be unfair to the patient to sensitize him to a useful agent. Penicillin topically is the most frequent drug producing sensitivity and probably should never be used unless it is the only agent to which the organism is sensitive. Various drugs containing mixtures of neomycin, bacitracin and polymyxin are therefore very popular among ophthalmologists. Sulfacetimide is also an excellent drug. The solution is 30% and the ointment 10% for the treatment of conjunctivitis. Because of its solubility, highly concentrated solutions may be prepared and, fortunately, these same solutions are associated with an extremely low incidence of sensitivity.

I feel that antibiotics such as aureomycin, acromycin and erythromycin should not be used in conjunctivitis because of the danger of sensitivity.

Moist heat applied to the closed lids is also helpful and makes the patient

* Assistant Clinical Professor, Department of Ophthalmology, The Chicago Medical School.

more comfortable. We advise the patient to apply the heat for about twenty minutes every three or four hours, using a clean wash cloth and plain warm water. We also advise him to use water about as warm as he would bathe in.

Acute purulent conjunctivitis is a mildly contagious disease. Therefore, it is advisable to suggest that the patient utilize a separate towel and wash cloth and soap during the course of the disease. Children should be kept home from school during the first twenty-four hours.

Ideally, it would be desirable to have smears and cultures to determine sensitivities on all cases of conjunctivitis. Practically, this is impossible. It is our practice to place patients on therapy when first seen and change drugs if the condition has not cleared within thirty-six to forty-eight hours. Only after seventy-two hours without improvement do we insist on identifying the organism and smears and cultures are taken. Admittedly, identification after therapy is difficult, but, realistically, this is the only practical approach.

Of late, it has become very popular to combine various antibacterial drugs with one of the steroids in a topical, ophthalmic preparation. It seems to me that the disadvantages of such a preparation outweigh the advantages when selecting a drug for the treatment of conjunctivitis. The only theoretical advantage is a more rapid paling of the eye. On the other hand, one is never sure without cultures and sensitivities that you have selected the proper antibacterial agent, and the steroid may only diminish the normal body reaction to infection, allowing it to spread and reach alarming proportions. Furthermore, virus and fungus infections, which are not uncommon in the eye, seem to thrive on topical steroids. These latter may mimic in their onset acute purulent conjunctivitis.

INTRODUCTION TO THE CONCEPT OF COMPREHENSIVE MEDICINE *

HARRY H. GARNER, M.D.** and ABRAHAM J. SIMON, Ph.D.***

Under the aegis of "scientific medicine," a primary and almost exclusive emphasis in medical education has been placed on the organic and chemical aspects of illness and treatment. Medical research became conceived largely as a laboratory process. The basic sciences were anatomy, physiology and chemistry. Illness became conceptualized as damage to, or deviation in, organ structure or as imbalance in body chemistry. The harmful effects of physical trauma, or of the intrusion or generation of noxious substance, seemed to be "the all" and "end all" of modern scientific medicine. A tendency toward deparmentalization, specialization, and technical manipulative procedures became characteristic of medical practice and medical education under this type of orientation. As a consequence, more and more has become known about less of the human body.

Illnesses are now diagnosable, preventable and curable in biochemical terms in our society in massive numbers. We take pride in this real achievement of "scientific medicine." Yet is is a curious contradiction that when medicine is enjoying its highest prestige for scientific knowledge and success with the organically ill person, there appears also significant criticism and a seriously impaired public confidence in the profession. While there may be many reasons for this, one of the factors must be the number and types of illnesses in which a strictly physiological-chemical orientation does not fit, cannot validly diagnose, and seems unable to heal. Also, there continue to flourish other healing arts and professions not so rigidly oriented to the values of the physical sciences and of scientific method which enjoy considerable public confidence and usage.

Both the lay public and many in the medical profession have expressed the need for a return to the good old days of the family doctor, who may have been short on science but was said to be long on intimate humaneness and sensitivity to everything his patient was experiencing. Perhaps there has developed in our kind of society, as many theoreticians and social philosophers believe, an increased incidence of troubled people generally, of psychosomatic disorders, of mental illness, and of social deviance. Certainly no one questions the reality of a greater preoccupation with and concern about such disorders. To whom other than the medical profesion can society turn for research, understanding and help with these concerns? Increased perception of such problems naturalistically as illness rather than moralistically or legally as badness retribution or inferiority tends to place more and more social responsibility for dealing with these ills on the doorstep of medicine.

Criticisms from within the medical profession have focused on the:

(1) Fragmentation of the patient into organ systems, and subordination of the doctor-patient relationship to the doctor-organ relationship. The doctor confines his interest to the afflicted organ or part of the patient, as though the doctors are dealing with that thing rather than with a person. The doctor treats the disease in a person rather than a person who has a disease.

(2) Limited knowledge about the idiopathic and functional disorders, about the relationship between psycho-social variables and illness, about these realistic and subjective stress situations and disturbances of mood, thought, and sensation behavior which impell people by the millions to call on physicians for help,

*Originally presented as a lecture to the Class of 1962, The Chicago Medical School, in the course of their orientation program, October 1958.

**Professor and Chairman, Department of Psychiatry and Neurology, The Chicago Medical School.

***Assistant Professor of Social Sciences, Department of Psychiatry and Neurology, The Chicago Medical School.

notwithstanding grossly inadequate evidence of biological or chemical pathology.

(3) Limited recognition that continuous exposure to such stress situations over a long enough period of time leads to organic breakdown, a phenomenon which complicates the problems of the physician in correctly diagnosing and then attempting to cure while the patient continues under the same stresses compounded by secondary stresses induced by the breakdown itself.

Limitation of the concept of illness to organic illness has become obviously unrealistic. Advances in public health, physiology itself, psychiatry and psychoanalysis, social work and the social sciences, esepecially sociology, anthropology and psychology, have all established etiological significance for linkages between psycho-social variables and illness. This in turn has led to the conception of holistic or comprehensive medicine as a hopeful approach to their diagnosis and treatment.

Recognition of the psycho-social components of illness, prevention and treatment is nothing new. This allegedly new interest in the patient's social situation, in the emotional, attitudinal and interpersonal stress components of illness has really never been denied. For a while it was ignored in favor of the less complex, more conspicuous and more objective etiological factors. Comprehensive Medicine now represents a re-emphasis and a re-application of old principles to the problems of diagnosis and treatment. There is though this one difference.

In the past, recognition and application of the psycho-social elements was largely based on such factors as personal interest, intuitive insight, life-long friendships or neighborliness between doctor and patient. It may have been part of the social life of the group, of the prevailing social customs as in the practice of folk medicine, of religion or faith healing or the communally respected intervention of the medicine man, the Shaman. Now recognition and application must meet the criteria of the more adequately developed behavioral sciences, based on tested experience, controlled observation, and, if possible, validated by the experimental procedures of the social sciences. Questions and answers about such illnesses would need to be tested as hypothesis by research in the social as well as the biological sciences.

It would be fair to say that medical education today is still to a large extent primarily structured for the teaching of traditional scientific medicine bio-chemically conceived, but that a vigorously active ferment is taking place to incorporate concepts and knowledge of the psycho-social factors contributing to illness and treatment as these are being currently experienced by physicians and patients. This activity we conceptualize as the teaching of integrated, holistic, or comprehensive medicine.

We emphasize understanding the whole person as an individual, a member of his family, his social class, his ethnic group, and as a citizen of his society and the relationship of these to his complaints, his illnesses and his treatment. The intervention of the physician then becomes another social variable, and a potentially controlling variable, making for health.

The process of transition from bio-chemical to comprehensive medicine is a low one. They key growing point of the process is the learning which takes place among medical students.

This process of incorporating a psycho-social with a bio-physical-chemical orientation seems particularly difficult for medical students, for reasons which are not yet quite clear. We are still in the beginning stages of this process, and in the coming years we expect psycho-social medicine to become inseparably integrated with bio-chemical medicine in the social role of the physician who is scientifically trained in both. The model physician would have to be both a healer and a humanist, comfortable in the physical as well as the social laboratory, and a productive responsible member of the community of which he would also be a student.

We aspire to stimulate the medical student to broaden and intensify his interest in all of man's activities, intellectual, physical and psycho-social, and to understand the human being as constantly bathed by his "external sea."

Sometimes, this is a soothing, or stimulating and invigorating experience, similar to that at a spa. More often, and especially for the sick, it is akin to a sea of troubled waters, which so engulf the individual that it is impossible for him alone to hold his head above the waves.

The ideal doctor must be as much aware of this psycho-social "bath" as he is of the internal cellular baths and their homeostatic balances. He must not allow idiosyncratic personal desires and strivings to so blind him that they interfere with the growth and spiritual well-being of his own self as well as that of the patient. In the light of our culture's stresses and strains, he must have insight into hopes and the frustrations, joys and the sorrows, and the many conflicts which patients, as people, endure. He must view patients with psychological X-ray vision also in full focus. It is not enough to have the Roentgen ray reveal the shadows of the spirit, as well as in the fossae of the skull, for the inner meaning of symptoms and for clues to the curative process.

"The ideal doctor," it has been said, "is one who understands the human body as completely as possible in the light of all the relevant sciences, who understands the human mind and its workings as well as is possible in the light of modern psychology and psychiatry, and who, in addition, understands a **human being for his full potential cultural and spiritual stature."** This image of the ideal doctor is first approached through the intellectual ability to grasp and retain many details of anatomy, chemistry, physiology, bacteriology, and other allied physical sciences. It is further approached through learning the clinical data of illness entities in the classroom and by experiencing them in the clinic and hospital working together with teachers and patients, those vital interpersonal relationships which, as Hippocrates said, "Impart by precept, lecture, and every other mode of instruction."

But lectures on medicine, surgery, obstetrics or pediatrics, and the examination of patients to ascertain whether the machinery of the body is well-ordered and geared for efficient work are not enough.

Essential to approaching the ideal is another qualification, an understanding of the human mind. To comprehend a person "in his full potential and spiritual nature" requires, first, an awareness of one's self as a human being. Our patients, in their dependence upon us, may wish us to don the garb of ruler or priest, or even to invest us with the omniscience and omnipotence of a diety the better to care for them. It is necessary to retain an awareness of self as just another person. To assist us in maintaining this perspective, and to understand the tie-in of the human body and mind, we have accepted as a goal of medical education, instruction also in the behavioral sciences and of human illness as a special form of behavioral deviation, which are functions of the stress of life.

To this end, knowledge of the patient is required as a functioning organism conducting many transactions also on interpersonal, social and psychological levels. How these have reciprocal, feedback connections with the sub-organ systems of the body will require knowledge of the person's patterns of behavior and value orientations, as well as of his illnesses such as is elicited by wider cross sectional and deeper longitudinal study through time.

The ways and means of attaining such a goal are available. But awareness of such a goal, and high premium on learning this method of practicing medicine is the necessary ingredient for the student physician and practitioner to bring to the learning process.

ESSAYS ON MEDICAL STATISTICS

II. The Reliability of Statistical Data

ERNEST B. ZEISLER, M.D.*

The function of statistical reasoning is to draw such conclusions as are warranted by the data. On this same principal are based the detection of a criminal, the trial of a defendant, and the diagnosis of disease. No intelligent physician will try to make a diagnosis without evidence; but even if the history, the physical findings, and the results of various laboratory tests and examinations by instruments of precision are all reliable, it may still be difficult to make the correct diagnosis. It is clear that if any of the evidence is faulty, the chance of correct diagnosis is further diminished. The prime requirement, therefore, is that the evidence should be reliable; in statistics this means that the data should be reliable.

The problem of the reliability of evidence is the same in statistics as in other fields, and depends upon many things. In the first place, the correctness of an observation depends upon the accuracy of one or more of an observer's senses, such as vision or hearing or touch; it may also depend on the accuracy of some laboratory test or of some chemical solution or some instrument of precision. Secondly, reliability depends upon the lapse of time between the act of observation and a record or a report made by the observer; in general, the longer the time the greater the opportunity for a lapse or a quirk of memory, and the less reliable the spoken or written report. There is an additional hazard because of the possibility of a slip of the tongue or of the pen and, in the case of a written report, the possibility of error in reading or in copying it.

All these factors can be affected by personal bias. It has repeatedly been observed that different witnesses to the same simple event may immediately thereafter give quite different and mutually inconsistent descriptions of it. Entirely innocent persons have often been convicted of serious crime because of mistaken identification by honest and well-intentioned witnesses; in twenty-seven of the sixty-five cases reported by Prof. Borchard in his classic book[1], there was mistaken identification of one or more defendants by one hundred and thirty-eight different witnesses. Judge Frank wrote[3.1]:

> Men are prone to see what they want to see.

This is true in science as well as in law. Frank said further[3.2]:

> It must be admitted that at the present day the testimony of even a truthful witness is much over-rated.

The same problem arises in the study of history. Prof. Gottschalk says[4.1]:

> The writer who thinks he has no philosophy of history or who believes he is detached is self-deceived, unless he is more than human, and therefore more likely to deceive others than if he were deliberately lying.

The effect of the unconscious workings of bias has long been recognized and often colorfully expressed. Tristam Shandy said[7]:

> It is the nature of an hypothesis, when once a man has conceived it, that it assimilates every thing to itself, as proper nourishment; and, from the first moment of your begetting it, it generally grows the stronger by everything you see, hear, read, or understand.

Sherlock Holmes put it this way[8]:

> It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.

* Clinical Associate Professor of Medicine, The Chicago Medical School.

There are various methods, none of them infallible, of attempting to minimize the chance of being misled by testimony which may have been colored by personal bias. Gottschalk says[4.3]:

> When the thought patterns and preconceptions of a witness are known and yet he states something out of keeping with them—in other words, if statements are *contrary to the witness' expectations or anticipations,* they have a high degree of credibility.

Another hint[4.1]:

> . . . an important general rule: *for each particular document, the process of establishing credibility should be separately undertaken regardless of the general credibility of the author.*

Two illustrations of the operation of personal bias are as follows. It is well-known that many illnesses, such as those due to the Coxackie viruses or the ECHO viruses, may be indistinguishable from paralytic poliomyelitis except by careful virus studies. A physician who sees a patient with such an illness will almost always inquire whether or not the patient has been vaccinated with the Salk vaccine, and the answer may well influence his diagnosis; if he is biased in favor of the vaccine, then he will tend to make some diagnosis other than poliomyelitis if the patient has been triply vaccinated, and will tend to make the diagnosis of poliomyelitis if the patient has not been vaccinated (or not triply vaccinated). In this way there will be an accumulation of new data in favor of the protective value of the vaccine. The National Foundation for Infantile Paralysis[5] stated that in 1956 a total of 188 cases of poliomyelitis were reported in the United States in persons who had been triply vaccinated with the Salk vaccine, and that 34 of these were paralytic. They said further that six of these 34 had been removed from the list after careful investigation, and that nine were still under investigation; the remaining 19 had been confirmed and had residual paralysis. They admitted that:

> More stringent criteria have been applied in establishing these triply-vaccinated cases as paralytic poliomyelitis than are used in the routine reporting of polio. For this reason it is not possible to directly compare these triply-vaccinated cases with other groups of paralytic cases, vaccinated or unvaccinated.

In other words, whenever paralytic poliomyelitis was reported in a triply-vaccinated person, special studies were undertaken which might lead to a change in diagnosis or to the discovery that the patient had not been triply vaccinated. But no such studies were instituted if paralytic poliomyelitis was reported in a person who gave no history of having been triply vaccinated, although such studies might have shown either that the disease was not really poliomyelitis or that the person had in fact been triply vaccinated. This shows obvious bias in favor of the vaccine. Further[5]:

> Three deaths from paralytic poliomyelitis have been reported in triply-vaccinated persons. One was well-established as bulbar poliomyelitis, but a re-check of vaccination records revealed that the child had never been vaccinated. A second case was clinically consistent with poliomyelitis; but pathological review revealed instead the anatomic findings of acute disseminated encephalomyelitis. The one remaining fatal case occurred in a five-year-old boy who died the day after onset of fever and within 30 minutes of hospitalization. An autopsy performed without examination of the brain or spinal cord was reported as poliomyelitis. This is the only fatal case now being carried in the registry.

But when a person was reported to have died of paralytic poliomyelitis and not to have been triply vaccinated, no special effort was made to rule out the possibility that he had been triply vaccinated, and no autopsies were uniformly performed in an attempt to change the diagnosis. The entire study has from the beginning contained enough bias to cast grave doubt on its conclusions.

In a comprehensive report on the use of anti-coagulants in what had been diagnosed as myocardial infarction, Wright, Marple and Beck admit that[9.2]:

Cerebral episodes constituted a diagnostic problem because of the difficulty of distinguishing cerebral emboli from cerebral hemorrhages and temporary functional disturbances. In the case of suspicious pulmonary symptoms, the problem was primarily that of distinguishing embolic phenomena from pneumonia.

It is quite clear that a physician biased in favor of the use of anticoagulants would tend to attribute cerebral or pulmonary disturbances to emboli rather than to hemorrhage, and that a physician biased against anticoagulants would tend to do the opposite. Inasmuch as [9.2]

> . . . the opinion of the physician who had attended the case was accepted as final for purposes of classification.***

any bias on the part of the physicians could play a considerable role, especially in view of the fact that a number of the participating physicians had already unequivocally demonstrated their bias in favor of the use of anticoagulants, as admitted by the author[9.1]. Although the authors themselves say that[9.3]

> . . . data in which any element of selection has entered are suspect***, they do not apply this criterion to their own report.

A piece of evidence is, in general, more reliable if it comes from two or more witnesses than if it comes from only one witness; this is the reason for the restriction in the Constitution of the United States against conviction of treason[2]:

> . . . no person shall be convicted of treason, unless on the testimony of two witnesses to the same overt act, or on confession in open court.

But testimony is not necessarily made more reliable by witnesses who have been influenced by the statement of another witness, that is by witnesses who are not independent. And two witnesses cannot be shown to be independent unless there is certainty that neither of them *can* know, when he gives his testimony, what the other witness has said or will say.

Testimony ceases to be reliable if a witness materially alters it after he has once given it, or if he gives testimony materially different from that which he has once given, and this is the basis of the eminently sensible decision of the Supreme Court of the United States in the Jencks case[6], in the effort to protect a defendant against the use of perjured testimony by the Government.

It is obvious that evidence which is self-contradictory cannot be valid and must be discarded at once. The trained statistician may be able to detect inconsistencies in statistical data which are not readily detectable by others. Suppose, for instance, we are presented with the following data: In a ward containing 60 patients there are

exactly 46 with heart disease
exactly 36 with hypertension
exactly 54 with arteriosclerosis
exactly 48 with renal disease
exactly 3 with all four of these

The mathematician can see at once that this is self-contradictory, for if the first four of these statements are true, then at least four patients have all four conditions, and the last statement cannot be true.

[**The criterion is as follows: Let $A_1, A_2, \ldots, A_m$ be distinct attributes, any one or more of which can be present in any number of a set of N individuals. Let (A_1) be the total number of individuals in the set in which A_1 is present, (A_2) the number in which A_2 is present, etc., and let $(A_1A_2 \ldots A_m)$ denote the number in which all m attributes are present simultaneously. Then,**

$$(A_1A_2 \ldots A_m) \geq (A_1)+(A_2)+\ldots+(A_m)-(m-1)N \quad]$$

Or suppose we are given the following data: In a ward containing 60 patients there are

exactly 36 with heart disease
excatly 25 with hypertension
exactly 18 with renal disease
exactly 11 with heart disease and hypertension
exactly 12 with heart disease and renal disease
exactly 6 with hypertension and renal disease
exactly 4 with all three
exactly 8 with none of the three

Again, the statistician can say quickly that if the first seven statements are true, then there must be exactly six patients who are free of the three diseases, so that the last statement is inconsistent with the others.

[The criterion is as follows: With the notation used above, and the further notation that (A_1A_2) denotes the number having both attributes A_1,A_2, etc., and $(A_1$ or A_2 or $A_3)$ denotes the number having at least one of the three attributes, then

$$(A_1 \text{ or } A_2 \text{ or } A_3) = (A_1)+(A_2)+(A_3)-(A_1A_2)-(A_1A_3)-(A_2A_3)+(A_1A_2A_3) \quad]$$

It is of great importance to distinguish sharply between *evidence* and *opinion*, even if the opinion is based on evidence. Law students learn this at an early stage of their study, and they find it important throughout their careers, for ordinarily only evidence is admissable in court. Although the distinction is fully as important in medicine, and especially in medical statistics, few physicians ever learn it. As a consequence, physicians often accept as evidence that which is mere opinion . The insidious effect of such confusion is shown in the previously mentioned studies concerning the Salk vaccine and the use of anticoagulants in myocardial infarction.

REFERENCES

1. Borchard, E. M.: *Convicting the Innocent.* Garden City Publishing Co., Inc., New York. 1932.
2. *Constitution of the United States,* Art. III, Sec-3.-1.

3.1 Frank, J. and Frank, B.: *Not Guilty.* p. 107. Doubleday & Co., Inc., Garden City, New York. 1957.

3.2 Frank, J. and Frank, B.: *ibid.,* p. 199.

4.1 Gottschalk, L.: *Understanding History. A Primer of Historical Method.* p. 9. Alfred A. Knopf, New York. 1950.

4.2 Gottschalk, L., *ibid.,* p. 143.

4.3 Gottschalk, L.: *ibid.,* p. 164.

5. *Information for Physicians on the Salk Poliomyelitis Vaccine.* No. 4, p. 28, February 1957. National Foundation for Infantile Paralysis, New York.
6. *Jencks vs. United States,* 353 U. S. 657.
7. Sterne, L.: *The Life and Opinions of Tristram Shandy,* Book II, Chap. XIX.
8. Watson, J. H.: *A Scandal in Bohemia,* in *The Adventures of Sherlock Holmes,* London, 1892.

9.1 Wright, I. S., Marple, C. D. and Beck, D. F.: *Myocardial Infarction. Its Clinical Manifestations and Treatment with Anticoagulents. A Study of 1031 Cases,* pp. 19-25. Grune & Stratton, New York. 1954.

9.2 Wright, I. S., Marple, C. D. and Beck, D. F.: *ibid.,* p. 194.

9.3 Wright, I. S., Marple, C. D. and Beck, D. F.: *ibid.,* p. 263.

ESSAYS ON MEDICAL STATISTICS

III. Random Sampling

ERNEST B. ZEISLER, M.D.*

A *universe* (or *population*) in Statistics is a class (or aggregate) of things which are for some purpose considered together[8]. One way of finding out what we may wish to know about this universe is to examine every one of its members, and then to organize the information in a suitable way; we thus obtain a description of the universe, and no statistical problem arises. If this method is not feasible, either because we do not have access to every member of the universe, so that it is impossible, or because the universe contains so many members that it is not practicable to examine them all, then we attempt to learn whatever we can by examining one or more than one *sample* of the universe, where by *sample* we mean a *proper part* (a part less than the whole).

It should be evident that no matter how large a sample is examined or how many samples are examined, if there is one member which is not examined *it is impossible to know with certainty* what might be found in the unexamined part of the universe. Suppose, for example, that from a bag known to contain 100 marbles we draw one marble at a time and do not replace it, continuing in this way until we have drawn 99 marbles, and suppose that all these 99 marbles are white. We may then consider it very "likely" or "probable" that the marble remaining in the bag is also white, but we cannot be certain of this, for the bag may have contained 99 white marbles and one black marble. Similarly, no matter how large a sample of the universe has been examined, there always remains some uncertainty about the unexamined part. Whatever conclusions we may draw from a sample about the entire universe must be expressed in terms which reflect this uncertainty.

The question immediately arises as to how it is possible to draw any conclusions at all, no matter how modest, about a universe so long as any of it has not been examined. The answer is that *there is no way, unless we have made certain assumptions* about the universe.

If we knew that a sample were completely representative of the universe, then it would follow by definition that whatever is true of the sample is also true of the entire universe, in proportion to the number of individual members of the sample and the universe. Although there is no way of knowing in advance that a given sample is completely representative of a given universe, we can at least attempt to proceed in such a way that there is no reason to exclude the possibility that the sample should be representative. Suppose, for instance, that we wish to learn something from a sample concerning a particular aspect of the entire population of the United States of America. We know in advance that no sample which contains less than 50 individuals can possibly be representative, for at least one State will not be represented in the sample; indeed, in order to be completely representative, the sample must contain the same proportions of people from the various States as the total populations of the States bear to each other; and, depending upon what we are studying, the sample may also have to contain enough individuals from each State to allow for various factors such as age, sex, race, and perhaps many other things. We can say that in general a sample should be large enough and varied enough to have represented in it all pertinent factors. But here there is already a rub, for one often does not and cannot know at the outset what all the pertinent factors are: there may be factors which are of considerable impor-

* Clinical Associate Professor of Medicine, The Chicago Medical School.

tance but of which we are totally unaware.

Besides being large enough and varied enough so that, as far as we can say, it *could* be representative, a sample must, as far as possible, be *random* in the sense that every member of the universe must have the same chance of being chosen to be in the sample as has every other member. And now there is a second difficulty, for unfortunately *there is no way of knowing with certainty that a given sample is truly random.*

The appearance of randomness is never conclusive evidence. This may be illustrated by the *Monge shuffle* (after Gaspard Monge, 1746-1818, the inventor of Descriptive Geometry): a deck of cards is shuffled by removing its top card to start a new pile; the second card of the deck is then placed on top of the first card, the third card of the original deck is placed below the new pile, and so on according to the rule that every odd card from the original deck is placed at the bottom of the new pile and every even card on top of the new pile, until the deck is all in the new pile. Inasmuch as the total number of different arrangements of the cards is some finite number N, it is evident that the cards must return to their original order in not more than N successive Monge shuffles. For a deck of 52 cards arranged by suits as in a new deck, eleven successive Monge shuffles bring the cards into an order which on cursory inspection appears to be random; yet one more Monge shuffle brings the cards into their original order.

[The general rule for a deck of 2n cards is that the original order is first restored after m Monge shuffles, where m is the smallest integer for which one of $2^m + 1$ and $2^m - 1$ is divisible by $4n + 1$. For a deck of 52 cards, 2n is 52, so that $4n + 1$ is 105, and the condition is that one of $2^m + 1$, $2^m - 1$ is divisible by 105. Now $2^{12} - 1 = 4095 = 39 \times 105$ so that the condition is satisfied for $m = 12$; it is easily seen that it is not satisfied for any smaller number.] It follows that a new deck of cards can be made to appear well-mixed by a single shuffle which is the reverse of the Monge: remove the cards one at a time alternately from the top and bottom of the deck, putting each card on top of the new pile at each stage. This illustrates the important fact that *it is theoretically impossible to prove that a given set of things is random;* the most that can be said is that no lack of randomness, that is no order, has been discovered.

It follows that it is never possible to be certain that a proposed method of choice will yield a truly random example, for it may involve an unknown source of randomness. We repeat that[8]:

> Plausible devices such as the blind or even mechanical mixing of marbles or beads, or the shuffling of dice or of cards, have been found to be inadequate to assure random sampling.

There is no mechanical device which may not have or develop a flaw such as to favor its operating in some ways more than in others; this, of course, is the basis on which a player at roulette has at times been able to break the bank. The same objection applies to every nonmechanical method of choice. Yule and Kendall give the following instructive examples[7.1]:

> It might be thought that any purely haphazard method of selection would give a random example. For example, if we wished to obtain a random sample of local tradesmen, one way which suggests itself is to take a Trades Directory, open it "at random" and take the first name on which the eye alights, repeating the process until the sample is of the required size. Or again, if we wished to obtain a random sample of wheat growing in a field, it might be thought that a satisfactory method would be to throw a hoop in the air "at random" and select all the plants over which it fell.

That such methods are apt to be deceptive may be seen from the two examples just given. In the first place, if we consulted a Trades Directory which had already been used, we should probably find that it opened at some pages more readily than at others; we should therefore tend to get the more popular tradesmen. More-

over, our eye might tend to be caught by long names or peculiar names. In either case some tradesmen would have a greater chance of being chosen than others, and the sample would not be random.

Again, in the second example, our hoop might tend to be caught by the taller ears of wheat, or we might tend unconsciously to throw it towards parts of the field where the wheat looked to be about the average height. These and other factors would destroy the random character of the sampling.

Again [7.2]:

> Sight is not the only sense which may bias a sampling method. In certain experiments counters of the same shape but of different colors were put in a bag and chosen one at a time, the counter chosen being put back and the bag thoroughly shaken before the next trial. On the face of it this appears to be a purely random method of drawing the counters. Nevertheless, there emerged a persistent bias against counters of one particular color. After careful investigation the only explanation seemed to be that these particular counters were slightly more greasy than the others, owing to peculiarities of pigment, and hence slipped through the sampler's fingers.

There is no doubt but that Yule and Kendall are justified in saying[7.4]:

> We can never be absolutely certain that a method of sampling is random. There are doubts on *a priori* grounds because for any given method there are always *conceivable* sources of bias, and we can never rule out entirely the possibility that some of these sources are present. The utmost we can do is to make their presence extremely unlikely by taking great care with the experiment.

Whether or not we can actually make the presence of sources of bias "extremely unlikely" is questionable, but at least we can attempt to minimize them, however large they may remain. One who does not understand the necessity for doing this, or is not acquainted with some of the less obvious sources of bias, is courting trouble if he essays to choose a random sample.

Frequently encountered but more or less obvious sources of bias are illustrated in what follows: (1) An increase in the number of recorded cases of crime or of some particular disease may be due merely to better reporting or to better recording or keeping of records. (2) The mortality rate in California in the 1930's was obviously not representative of the mortality rate throughout the United States, for many old or sick people moved to California to spend their last years. (3) Hospital data cannot be regarded as representative of the community because hospital population is clearly not a random sample: persons who are ill are *ipso facto* not a random sample, and persons who are seriously ill are more likely to enter the hospital than those less seriously ill. (4) No sample consisting of volunteers for some procedure (*e.g.*, vaccination) is a random sample, for the very fact that people volunteer indicates a difference between them and people who do not volunteer, and this difference may affect the results of an experiment. Hill says[1.1]:

> A sample which is composed of volunteers or self-selected individuals is not likely to be a random sample of the universe from which it is drawn. If, for example, the prevention of colds by vaccine is offered to a group of persons, the volunteers may belong mainly to that section of the group which suffers most severely from colds and hopes for some advantage from the treatment. They are in that event a select group, not comparable with the remainder of the population from which they were drawn. In such cases the question must always arise: Is the act of volunteering correlated with any factor which may have an influence upon the final results of the experiment?

Again[1.1]:

> Mothers who bring their babies to be inoculated against an infectious disease may be the more intelligent who take more and better care of their children. Also, such mothers may be more frequently the mothers of single children or of a small family; . . . A

comparison of the inoculated volunteers with the uninoculated non-volunteers, therefore, involves important differences between the groups apart from the state of inoculation, differences which may well influence the relative incidence rates of disease and thereby produce quite misleading comparisons. Without very good evidence of equality such comparisons of volunteers and non-volunteers should not be made.

(5) Inquiry by questionnaire is subject to serious error.. As Hill says[1.2]:

> Inquiries carried out by means of questionnaires are *par excellence* those in which selection must be suspected. In all such inquiries replies to the questions put — even to the simplest question — are received from only a proportion of the individuals to whom the form is sent. There can never be the slightest certainty that the individuals who choose to reply are a representative sample of all the individuals approached; indeed very often it is extremely unlikely that they are representative.

A well-known example of the unreliability of the questionnaire method is given by Levinson[4], though here there was also grievous error in selecting the group to which the questionnaire was to be sent:

> One of the most notorious examples of an investigation based on a biased sample is the *Literary Digest* poll of voters in the election of November 2, 1936. The interest and excitement over this poll were due to the fact that it was by far the largest of several political polls, ballots being mailed to ten million voters, with over two million responses, and to the fact that the *Literary Digest* had been making similar polls for nearly twenty-five years and had never failed to predict correctly the result of an election.
>
> * * * * *
>
> The fact is that the poll itself contained convincing evidence that something was radically wrong, and this evidence was seen and interpreted at the time by many competent statisticians. The nature of this evidence is as follows: The completed poll carried figures showing "How the same voters voted in the 1932 election." These figures showed that 1,078,012 of those who sent their ballots to the *Digest* poll voted the Republican ticket in 1932, and 1,020,010 the Democratic ticket. But this is definite proof of a biased sample, for it actually shows more Republican 1932 voters that Democratic, while the 1932 election itself showed 22,821,513 Democratic votes against 15,761,787 Republican votes.
>
> In other words, in 1932 there were about 45 per cent more Democratic voters than Republican voters . . .

Levinson says that a fair sample taken in 1936 should have contained about 45 per cent more persons who had voted Democratic than Republican in 1932. The bias in the sample is evident, and could have been due to choosing the names of the persons to whom ballots should be sent from telephone directories; telephone subscribers are a selected group, differing from non-subscribers in the important fact that the former are likely to be better-off financially than the latter, many of whom cannot pay for private telephone service. In any case, the woefully erroneous prediction of an easy victory for Landon cost the *Literary Digest* its life.

Many sources of bias are less obvious but not necessarily less invidious. One such is illustrated by Hill[1.3]:

> An interesting example of selection in taking a random sample of houses is suggested in the Ministry of Health's report on the influenza pandemic of 1918. (. . .) To obtain facts as to the incidence and fatality from influenza in 1918-1919 a house-to-house inquiry was undertaken in five areas of Leicester, information being obtained *so far as possible* at every fifth house. Houses which were found closed at the time of visit had to be ignored in this census. But houses in which there are young children are less often found closed and this would tend to affect the age-distribution of the population recorded in the sample. Compared with the population from which it was drawn the sample would be likely to contain an undue proportion of young

children and a deficit in the number of adults.

Enough has been said to show that the less random a sample the less reliable the conclusions based on it. Inasmuch as there is no way of being certain that a given sample is truly random, it follows that *no conclusion based upon sampling is completely reliable*, so that *no statistical conclusion can be absolutely certain*. But if we minimize the sources of bias in sampling, we maximize the likelihood that the conclusions drawn from the samples are correct.

Of all presently known methods of trying to attain randomness in sampling, the most successful is the use of what are called *Random Sampling Numbers*. The first large set of such numbers was compiled by Tippet[6], who "haphazardly" chose 41,600 digits from census reports and arranged them into 10,400 numbers ranging from 0000 to 9999, in the form of tables. Here, as elsewhere, there could be no certainty that the "haphazard" choice was actually haphazard, or that the choice of numbers and their arrangement were really random; the most that could be said was what Yule and Kendall did say[7,3]:

> Tippet's numbers have been subjected to numerous investigations which make their randomness for many practical cases highly probable.

And this was all Tippet claimed for them. Since that time they have been found inadequate in several lengthy trials[5]. A larger and more satisfactory set of random sampling numbers has been given by Kendall and Smith[3], who mechanically produced a series of 100,000 digits, which they then arranged in numbers of four digits, each such number consisting of two pairs of digits; these were all displayed in 100 successive tables of 1000 digits each; each of these 100 tables is made up of 25 rows of 40 digits each. For illustration, the first two rows of the first table of 1000 digits are as follows:

	1-4	5-8	9-12	13-16	17-20	21-24	25-28	29-32	33-36	37-40
1	23 15	75 48	59 01	83 72	59 93	76 24	97 08	86 95	23 03	67 44
2	05 54	35 50	43 10	53 74	35 08	90 61	18 37	44 10	96 22	13 43

Here also there can be no certainty that these numbers and tables are really random. They have, however, been examined for randomness by the following four tests[2]:

The frequency test. The number of times each digit occurs in any given block of the numbers is compared with its expected value, which is one-tenth of the total number of digits in the block.

The serial test. This is the frequency test applied to the pairs of digits, that is to the numbers from 00 to 99.

The poker test. Among the numbers of four digits each (e.g., 2315, 7548, 5901, etc., in the first row of the first table) one counts the number of instances of four-of-a-kind (e.g., 0000, 1111, 2222, etc.), the number of instances of three-of-a-kind, two pairs, one pair, and four different digits, and compares these with the expected values.

The gap test. The gaps (the numbers of digits between successive zeros, reading as in a book, from left to right in successive rows) are determined and their frequencies compared with expectation.

These tests were applied to the entire set of 100,000 digits, the four sets of 25,000 each in order, and to the twenty sets of 5,000 in order, and all the tests were satisfied by all these sets. When applied to the successive sets of 1,000 digits each, it was found that five of them failed in one or two tests, namely:

thousand	tests failed
26th	serial
47th	poker
49th	frequency, serial
81st	serial
90th	serial

Concerning these failures Kendall justly says[2]:

> One of the peculiar features of a table of Random Sampling Numbers is that it must contain here and there patches which are not suitable for drawing a sample when used by themselves. To put it roughly, the whole table should obey the tests, and so should most of its parts; but if the table when used as a whole is to supply a random sample or a series of random samples it must give the un-

usual a chance of occurring in its due proportion, and that part of the table which does so may be unsuited for use by itself.

It is the case that *all random sampling methods leak,* and great care must be taken in the use of even the best. In using Kendall and Smith's numbers, none of the five sets of 1,000 each which failed to pass all the tests must ever be used by itself, as has been said, though it may be used as part of a set of 5,000 or more. One may start anywhere and[2] read across like an ordinary page of print. This is the order in which they have been read to be tested. The tests are, however, equally valid if the numbers are read backwards and this procedure may also be adopted. I think it very unlikely that any bias would be introduced if the numbers were read in other ways, e.g. downwards, but it is as well not to incur the risk, however slight it may be.

If more than one sample is sought, then one must start using the numbers more than once, and should avoid, as far as possible, permitting any bias as to the starting point; the safest way to do this—and in this way all *personal* bias is eliminated—is to start each time with the digit which in the table immediately succeeds the last digit used in the last preceding time the numbers were used. Furthermore, once a scheme of sampling has been adopted for a particular kind of sample, this should not be changed for choosing subsequent samples of the same kind, unless a definite rule for changing it in certain ways and at certain intervals has been accepted before the first sample is chosen; this also is for the purpose of avoiding *personal* bias.

The novice is advised against attempting to use the table for random sampling without the assistance of someone experienced in its use. Nevertheless, we include a brief example of how it is to be used. Suppose we wish to choose a random sample of ten of fifty patients in beds numbered from 1 to 50. We start at the beginning of the table of numbers and read the digits in pairs, from left to right and in successive rows from top to bottom, as in a book; we choose the first ten numbers which do not exceed 50, and these are 23, 15, 48, 1, 24, 8, 3, 44, 5, 50. Thus the sample will consist of the patients in the beds numbered 1, 3, 5, 8, 15, 23, 24, 44, 48, 50. (Hill[1.4] gives a table of 10,240 digits from the numbers of Kendall and Smith, and these are arranged in sixteen lists of 640 digits each; there is no indication as to whether or not these shorter tables have been subjected to the four tests given above.)

Summary

(1) Every inference to a universe from a sample depends upon assumptions concerning the universe.

(2) An inference from a sample is more reliable the more nearly random the sample.

(3) Randomness in sampling is maximized by the proper use of Random Sampling Numbers.

(4) There is no way of knowing with certainty that a given sample is truly random.

(5) No conclusion based upon sampling is completely reliable.

REFERENCES

1.1 Hill, A. B., *Principles of Medical Statistics,* p. 18. Oxford University Press, New York, 1955.
1.2 Hill, A. B., *ibid,* p. 20.
1.3 Hill, A. B., *ibid.,* p. 23.
1.4 Hill, A. B., *ibid.,* p. 291.
2. Kendall, M. G., in Kendall, M. G. and Smith, B. B., *Tracts for Computers,* ed. by E. S. Pearson. No. XXIV. *Tables of Random Sampling Numbers,* p. vii, Cambridge University Press, 1946.
3. Kendall, M. G. and Smith, B. B., *op. cit.* in 2.
4. Levinson, H. C., *Your Chance to Win: The Laws of Chance and Probability,* p. 267, Farrar & Rinehart, Inc., New York, 1939.
5. Pearson, E. S., *op. cit.* in 2, p. v.
6. Tippet, L. H. C., *Tracts for Computers,* ed. by Karl Pearson, No. XV. *Random Sampling Numbers.* Cambridge University Press, London, 1927.
7.1 Yule, G. U. and Kendall, M. G., *An Introduction to the Theory of Statistics,* p. 337. Charles Griffin & Co., Ltd., London, 1940.
7.2 Yule, G. U. and Kendall, M. G., *ibid.,* p. 339.
7.3 Yule, G. U and Kendall, M. G., *ibid.,* p. 341.
7.4 Yule, G. U. and Kendall, M. G., *ibid.,* p. 346.
8. Zeisler, E. B., *Essays on Medical Statistics.* I. Introduction. The Quarterly, p. 00, 1958.

ESSAYS ON MEDICAL STATISTICS

IV. Statistical Association

ERNEST B. ZEISLER, M.D.*

It was stated in a previous essay[1] that the principles of statistical inference cannot be understood unless one knows the meaning of a number of statistical terms. In the present essay two of these terms, **statistical association** and **dissociation,** are considered in some detail, and their chief applications in medicine are outlined.

1. THEORY

Consider a universe of N individuals. Let A, B denote two attributes, either or both of which may be present or absent in one or more members of the universe, and let $\overline{A}$, $\overline{B}$ denote the absence of A, B, respectively. An individual having both attributes A, B will be said to be of **type** AB, one having A but not B of type AB, and so on. Let [AB], [$A\overline{B}$], [$\overline{A}B$], [$\overline{A}\overline{B}$] denote the number of individuals of types AB, $A\overline{B}$, $\overline{A}B$, $\overline{A}\overline{B}$, respectively; if we let these numbers be denoted by a, b, c, d, the distribution of individuals among the four types may be indicated by the array

(1)

	B	$\overline{B}$
A	a	b
$\overline{A}$	c	d

and each of a, b, c, d is an integer $\geqq 0$. The number [A] of individuals with attribute A is seen to be a+b, the number [$\overline{A}$] without attribute A (that is, with attribute $\overline{A}$) is c+d, and so on, as in the array

(2)

	B	$\overline{B}$	
A	a	b	a+b
$\overline{A}$	c	d	c+d
	a+c	b+d	N

DEFINITION 1. A is **independent of** B when A is present in the same proportion of individuals with B as in those without B.

It is seen by (2) that the proportion of individuals with B in which A is present is [AB] ÷ [B], which is meaningless if [B] is zero; we will, therefore, consider only cases in which [B] $\neq 0$, that is a+c $\neq 0$, and then the proportion is

$$\frac{[AB]}{[B]} = \frac{a}{a+c} \tag{3}$$

Similarly, the proportion of individuals without B in which A is present is [$A\overline{B}$] ÷ [$\overline{B}$], which is meaningless if [$\overline{B}$] is zero; hence we will consider only cases in which [$\overline{B}$] $\neq 0$, that is b+d $\neq 0$, and then the proportion is

$$\frac{[A\overline{B}]}{[\overline{B}]} = \frac{b}{b+d} \tag{4}$$

It now follows from Def. 1 by (3) and (4) that A is independent of B if and only if

$$\frac{a}{a+c} = \frac{b}{b+d}$$

that is, a(b+d) = (a+c)b, or ab+ad = ab+bc, or ad = bc, so that

THEOREM 1. A **is independent of** B **if and only if** ad−bc = 0.

If we interchange A and B in Def. 1, we see that B is independent of A when B is present in the same proportion of individuals with A as in those without A, that is

$$\frac{[AB]}{[A]} = \frac{[\overline{A}B]}{[\overline{A}]} \tag{5}$$

and these proportions are meaningless unless [A] $\neq 0$ and [$\overline{A}$] $\neq 0$, that is a+b $\neq 0$ and c+d $\neq 0$; with our previous restrictions, we limit ourselves to cases

* Clinical Associate Professor of Medicine, The Chicago Medical School.

in which $a+b \neq 0$, $c+d \neq 0$, $a+c \neq 0$, $b+d \neq 0$; since each of a, b, c, d is $\neq 0$, it follows that

$$(6) \quad a+b>0, c+d>0, a+c>0, b+d>0$$

It now follows by (5) that B is independent of A if and only if

$$\frac{a}{a+b} = \frac{c}{c+d}$$

that is, $a(c+d) = (a+b)c$, or $ac+ad = ac+bc$, or $ad = bc$, so that

THEOREM 2. **If A is independent of B, then B is independent of A.**

DEFINITION 2. A, B are **independent** when they are independent of each other.

And it now follows by Thm. 1, 2 that

THEOREM 3. A, B **are independent if and only if** $ad-bc = 0$.

DEFINITION 3. A is **positively associated with** B when A is present in a greater proportion of individuals with B than in those without B.

It follows from (3) and (4) that A is positively associated with B if and only if

$$\frac{a}{a+c} > \frac{b}{b+d}$$

that is, $ad > bc$, so that

THEOREM 4. A **is positively associated with** B **if and only if** $ad-bc > 0$.

It is easily shown that

THEOREM 5. **If A is positively associated with** B, **then** B **is positively associated** with A.

DEFINITION 4. A, B are **positively associated** when each is positively associated with the other.

And now it follows from Thm. 4, 5 that

THEOREM 6. A, B **are positively associated if and only if** $ad-bc > 0$.

DEFINITION 5. A is **negatively associated with** B when A is present in a smaller proportion of individuals with B than in those without B.

DEFINITION 6. A, B are **negatively associated** when each is negatively associated with the other.

It is easily shown that

THEOREM 7. **If A is negatively associated with** B, **then** B **is negatively associated** with A.

and that

THEOREM 8. A, B **are negatively associated if and only if** $ad-bc < 0$.

It is seen from Def. 5 that A is negatively associated with B when A is present in a greater proportion of those without B (that is, those with $\overline{B}$) than in those with B (that is, those without $\overline{B}$), so that

THEOREM 9. A **is negatively associated with** B **if and only if** A **is positively** associated with $\overline{B}$.

Similarly,

THEOREM 10. A, B **are negatively associated if and only if** A, $\overline{B}$ **are positively associated, and if and only if** $\overline{A}$, B **are positively associated.**

Furthermore,

THEOREM 11. A, B **are positively associated if and only if** $\overline{A}$, $\overline{B}$ **are positively associaited.**

and,

THEOREM 12. A, B **are independent if and only if** $\overline{A}$, $\overline{B}$ **are independent.**

DEFINITION 7. A **is completely associated with** B when A is not present without B, that is, when $[A\overline{B}] = 0$.

It is seen by (1) that

THEOREM 13. A **is completely associated with** B **if and only if** $b = 0$.

By interchanging A and B in Def. 7 it follows that

THEOREM 14. B **is completely associated with** A **if and only if** $c = 0$.

DEFINITION 8. A, B **are completely associated** when at least one of them is completely associated with the other.

It now follows by Thm. 13, 14 that A, B are completely associated when at least one of b, c is zero, that is when $bc = 0$; hence

THEOREM 15. A, B **are completely associated if and only if** $bc = 0$.

It follows from (6) that if either of b, c is zero, then $a > 0$ and $d > 0$, so that $ad > 0$; hence, if $bc = 0$, then $ad-bc > 0$, so that, by Thm. 6,

THEOREM 16. **If A, B are completely associated they are positively associated.**

DEFINITION 9. A, B are **equivalent** when they are completely associated with each other.

And now by Thm. 13, 14,

THEOREM 17. A. B **are equivalent if and only if** $b = c = 0$.

DEFINITION 10. A **is dissociated from** B when it is positively associated with $\overline{B}$.

It follows by Thm. 9 that

THEOREM 18. A **is dissociated from** B **if and only if it is negatively associated with** B.

and now by Thm. 7,

THEOREM 19. **If** A **is dissociated from** B, **then** B **is dissociated from** A.

DEFINITION 11. A, B **are dissociated** when they are dissociated from each other.

so that by Def. 10,

THEOREM 20. A, B **are dissociated if and only if each is positively associated with the negative of the other.**

DEFINITION 12. A **is completely dissociated from** B when it is completely associated with $\overline{B}$.

and now by Def. 7,

THEOREM 21. A **is completely dissociated from** B **if and only if** $[AB] = 0$, **that is** $a = 0$.

By interchanging A and B it is seen that

THEOREM 22. B **is completely dissociated from** A **if and only if** $a = 0$.

and

THEOREM 23. **If** A **is completely dissociated from** B, **then** B **is completely dissociated from** A.

DEFINITION 13. A, B **are completely dissociated when** A, $\overline{B}$ **are completely associated.**

If, without changing the entries in the array (1), we write it in the form

(7)

	$\overline{B}$	B
A	b	a
$\overline{A}$	d	c

if follows by Def. 13 and Thm. 15 that

THEOREM 24. A, B **are completely dissociated if and only if** $ad = 0$.

Hence,

THEOREM 25. A, B **are completely dissociated if and only if at least one of** $[AB]$, $[\overline{AB}]$ **is zero.**

It is seen by (6) that if at least one of a, d is zero, then $b > 0$ and $c > 0$, so that $bc > 0$, whence $ad - bc < 0$; hence, by Thm. 8,

THEOREM 26. **If** A, B **are completely dissociated they are negatively associated.**

and now by Def. 6, Thm. 18 and Def. 11,

THEOREM 27. **If** A, B **are completely dissociated they are dissociated.**

DEFINITION 14. A, B are **contradictory** when A, $\overline{B}$ are equivalent.

It now follows by (7) and Thm. 17 that

THEOREM 28. A, B **are contradictory if and only if** $a = d = 0$.

and it follows by (7) and Thm. 3 that

THEOREM 29. A, B **are independent if and only if** A, $\overline{B}$ **are independent.**

We can summarize what has been demonstrated as follows:

(8)

a, b, c, d	A, B	A, $\overline{B}$
$ad - bc = 0$	independent	independent
$ad - bc > 0$	positively associated	negatively associated (dissoc.)
$bc = 0$	completely associated	completely dissociated
$b = c = 0$	equivalent	contradictory
$ad - bc < 0$	negatively associated (dissoc.)	positively associated
$ad = 0$	completely dissociated	completely associated
$a = d = 0$	contradictory	equivalent

DEFINITION 15. The **coefficient of association** between A, B is

$$C = 50\left(\frac{ad-bc}{ad+bc} - \frac{b}{a+b} - \frac{c}{c+d} + 1\right) \tag{9}$$

which is admissible because $a+b \neq 0$ and $c+d \neq 0$ by (6), and we have seen that at least one of ad, bc is $\neq 0$. We may write (9) also in the form

$$(10) \quad C = 50 \left[\frac{ad-bc}{ad+bc} + \frac{ad-bc}{(a+b)(c+d)} \right]$$

It is readily shown that the coefficient of association can take on all values from -100 to $+100$, and that it varies as follows:

(11)

A, B	C
independent	0
positively associated	> 0
completely associated	$50 < C \leqq 100$
equivalent	100
negatively associated	< 0
completely dissociated	$-100 \leqq C < -50$
contradictory	-100

The definition here given for the coefficient of association differs materially from that chosen as most generally useful by Yule and Kendall[3]; the reasons for the present choice will be stated subsequently.

2. APPLICATIONS TO MEDICINE

There are three chief applications to medicine of the theory of statistical association: first, the question of the independence or association of two attributes, at least one of which is either physiological or pathological; secondly, the evaluation of a symptom or a sign or a laboratory test, or a particular set of such things, in the diagnosis of disease; thirdly, the evaluation of a prophylactic measure in the prevention of a given disease, or of a therapeutic measure in the treatment of a given disease. The clinician is interested especially in the second and third of these applications.

The Independence or Association of Two Attributes.

Suppose we wish to determine whether two things, say berry aneurisms at the base of the brain and congenital polycystic renal disease, are associated. If we denote these conditions by A, B, respectively, it is seen by (1) that **four** numbers are required: we must know the number **a** of persons having both berry aneurisms and congenital polycystic renal disease, the number **b** of persons having berry aneurisms but no congenital polycystic renal disease, the number **c** of persons having congenital polycystic renal disease but no berry aneurisms, and the number **d** of persons free of both. Only then can we determine the coefficient of association defined in (9). One may make the point that if it has been found that, say, **b** is zero, then we need not know **c**, for we already know that A, B are completely associated, as seen in (8); but this means only that A is completely associated with B (cf. Thm. 13), and does not give us the information necessary to determine the degree of association of B with A.

The same considerations apply to the study of the relationships between hygiene and tuberculosis, of cigarette smoking and bronchogenic carcinoma, of sex and coronary atherosclerosis. But great care must be exercised in many instances. For example, in the study of sex and coronary atherosclerosis, we may let A denote "male" and B denote "atherosclerosis of the coronary arteries," but this is quite inadequate. Inasmuch as the development of coronary atherosclerosis is never instantaneous, we will have to restrict A to mean males of at least a certain age, and $\overline{A}$ to females of at least the same age; furthermore, B will have to be restricted to atherosclerosis of at least a given degree. A more satisfactory way of studying such complex relationships is not by statistical association but by statistical **correlation** (to be discussed in a subsequent paper); in the given example, this would consist in considering various age groups and various degrees of coronary atherosclerosis. Here, as everywhere in statistics, we should not expect to reach sensible conclusions by the application of formulae by rote.

Another problem is that of the relationship between "Negro" and sickle-cell hemoglobin. But one must first agree on exactly what is meant by the term "Negro"; for example, we must agree on whether or not it is to include a person who is one-eighth Negro. Secondly, the presence of sickle-cell hemoglobin must not itself be used in deciding whether its possessor is or is not at least part Negro,

for otherwise there would be a **petitio principii,** which is always to be excluded from science.

Diagnostic Criteria.

Let A denote some particular symptom or physical sign, or some given outcome of a laboratory test or of a test by an instrument of precision, or a particular combination of such things, and let B denote a particular disease or some particular set of diseases, as in the following:

A	B
+ blood Kahn	syphilis
Argyll-Robertson pupil	tabes dorsalis
gastric anacidity	primary gastric carcinoma
L.E. phenomenon	disseminated lupus erythematosis
I^{131} uptake $> 50\%$	thyrotoxicosis

We wish to evaluate A as a possible criterion for the diagnosis of B. To this end we examine a number of individuals for the presence or absence of B, and the same individuals for the presence or absence of A. The first and **essential** restriction on our procedure is that in deciding whether or not B is present in an individual we must not use A itself, that is, **we must not use in the diagnosis of B the very attribute which we are trying to evaluate** as a criterion in the diagnosis of B. To proceed otherwise would be to commit a **petitio principii,** analogous to that previously mentioned. That this is not an idle warning is demonstrated by the fact that eminent scientists have unwittingly fallen into the trap. Darwin's principle[2] of "the survival of the fittest" serves as an example, for the criterion of fitness to survive was nothing other than the fact of survival, so that this famous principle amounted only to the tautology that "the survivors are those who survive." As meticulous a reasoner as Carlson[1] wrote as follows concerning addiction to alcohol:

> If the addict really desires to get over the addiction, then almost any therapy, or none at all, is successful in getting him over the addiction.

But this says exactly nothing unless we possess some criterion as to whether or not the addict "really" desired to get over the addiction, other than whether or not he did in fact get over it. To say that he "really" desired to get over it if he does get over it, and that he did not "really" desire to get over it if he does not, regardless of whether he did or did not say that he desired it, is to beg the question.

If one accepts as true the proposition that the Hargraves phenomenon is proof of disseminated lupus erythematosis, then one may thereafter use this criterion to make the diagnosis of the disease in every case in which the Hargraves phenomenon is elicited and refuse to make it in all other cases; in this way any error in the original assumption will be perpetuated. Or suppose one accepts the statement that multiple sclerosis is never cured; then no case of what appears in every other respect to be multiple sclerosis will ever be counted as multiple sclerosis if the patient is in fact cured, and whatever error there may be in the original statement will be perpetuated and, indeed, strengthened, because it will appear to have been confirmed in a larger number of instances.

Let us now return to our attempt to evaluate A as a criterion for the diagnosis of B. After having examined a number of individuals for B (**without** using A as a criterion for the presence of B) and also for A, we arrange our observations as in the array (2):

(12)

	B	$\overline{B}$	
A	a	b	a+b
$\overline{A}$	c	d	c+d
	a+c	b+d	

Now suppose the individuals in this completely examined universe are shuffled in such a way that we no longer know to which type (AB, $A\overline{B}$, etc.) any particular individual belongs. We then

ask what, if any, value A has as a test for B in **this universe;** this means that if we examine any one individual of this universe and find that he has the attribute A (or does not have it), then what, if anything, can we infer as to the presence or absence of B in that individual?

It should be evident that if A is present in all members of the universe, in those without B as well as in those with B, then A can have no value at all as a test for B, inasmuch as the presence of A in a given individual warrants no interference as to the presence or absence of B in that individual; the same is true if A is absent from every individual in the universe. Hence we limit our discussion to cases in which $[A] \neq 0$ and $[\overline{A}] \neq 0$. Similarly, if B is present (or is absent) in all members of the universe, then no test for B is needed, so that we limit ourselves to cases in which $[B] \neq 0$ and $[\overline{B}] \neq 0$. Hence we limit ourselves to cases in which $a+b$, $c+d$, $a+c$, $b+d$ are all different from zero; since each of a, b, c, d is $\geqq 0$, it follows that

$$(13)\quad a+b>0,\ c+d>0,\ a+c>0,\ b+d>0$$

which is the same as (6). Those $b+d$ individuals in whom B is not present, that is **those who do not have the disease** B, are called the **controls.**

But what we are actually interested in is not the value of A as a criterion for the presence of B in a universe each member of which has already been effectively examined for the presence or absence of B, but the value, if any, of A as a criterion for the diagnosis of B in individuals in whom the presence or absence of B has **not** been determined. And to this end **we must make certain assumptions** as to the distribution of A, B in an **unexamined** universe.

DEFINITION 16. A is **positive** when it is present and is **negative** when it is absent.

It follows from (12) that

$$(14)\quad \begin{aligned} &A \text{ is } + \text{ in } \frac{a}{a+c} \text{ of the individuals with B}\\ &A \text{ is } - \text{ in } \frac{c}{a+c} \text{ of the individuals with B}\\ &A \text{ is } + \text{ in } \frac{b}{b+d} \text{ of the individuals without B}\\ &A \text{ is } - \text{ in } \frac{d}{b+d} \text{ of the individuals without B} \end{aligned}$$

DEFINITION 17. A **is falsely positive** when it is positive in the absence of B.

DEFINITION 18. A **is falsely negative** when it is negative in the presence of B.

It now follows from (14) that

$$(15)\quad \begin{aligned} &A \text{ is falsely } + \text{ in } \frac{b}{b+d} \text{ of the individuals without B}\\ &A \text{ is falsely } - \text{ in } \frac{c}{a+c} \text{ of the individuals with B} \end{aligned}$$

On the other hand,

$$(16)\quad \begin{aligned} &A \text{ is falsely } + \text{ in } \frac{b}{a+b} \text{ of the individuals in which A is } +\\ &A \text{ is falsely } - \text{ in } \frac{c}{c+d} \text{ of the individuals in which A is } - \end{aligned}$$

One often speaks of "percentage" without making clear of **what** it is the percentage. In the present case, by the percentage of false positives (falsely positive tests) one could mean either percentage of individuals without B or else percentage of all positives, and these are, respectively,

$$100\frac{b}{b+d} \text{ and } 100\frac{b}{a+b}$$

which are not at all the same thing; they are, in fact, equal if and only if either $b=0$ or $a=d$. Similarly, the percentage of false negatives (falsely negative tests) could mean either percentage of individuals with B or else percentage of all negatives, and these are, respectively

$$100\frac{c}{a+c} \text{ and } 100\frac{c}{c+d}$$

which are not the same thing and are equal if and only if either $c=0$ or $a=d$. In order to avoid confusion it should always be clear exactly what we mean whenever we use the term "percentage." In the present case we wish to infer from the presence (or absence) of A as much as possible as to the presence (or absence) of B in the same individual; consequently, the more useful of the percentages are the percentage of false positives among all the positives and the percentage of false negatives among all the negatives; we will, therefore, use these terms hereafter only as follows:

DEFINITION 19. The **percentage of false positives** (or **negatives**) is the percentage among all the positives (or negatives).

Thus,

(17)

The percentage of false positives is $\frac{100b}{a+b}$

The percentage of false negatives is $\frac{100c}{c+d}$

DEFINITION 20. A is a **completely reliable test for** B when the presence of A implies the presence of B and the absence of A implies the absence of B.

Thus A is a completely reliable test for B if and only if there are no false positives and no false negatives, that is by (17), if and only if $b = c = 0$. And now, by Thm. 17,

THEOREM 30. A **is a completely reliable test for** B **if and only if** A, B **are equivalent.**

It is true that if $a = d = 0$, then the presence of A implies the absence of B and the absence of A implies the presence of B, so that whatever the outcome of the test for A, we can infer with certainty whether or not B is present. In this case we will say, in accordance with Def. 20, that A is a completely reliable test for $\overline{B}$, rather than for B. It follows by Thm. 28 that

THEOREM 31. A **is a completely reliable test for** $\overline{B}$ **if and only if** A, B **are contradictory.**

It may be said that completely reliable tests are relatively rare in clinical medicine.

The **usefulness** of a test A for a disease B depends on the purpose for which it is to be used. If our purpose is to be able to infer with certainty the presence of B from the fact that A is positive, then we need a test with no false positives; if our purpose is to be able to infer with certainty the absence of B from the fact that A is negative, then we need a test with no false negatives.

DEFINITION 21. A **is pathognomonic of (completely specific for)** B when there are no false positives.

Hence A is pathognomonic of B if and only if $[AB] = 0$, that is $b = 0$, from which it follows by Thm. 13 that,

THEOREM 32. A **is pathognomonic of** B **if and only if** A **is completely associated with** B.

If b is not zero but is relatively small, then A is not pathognomonic of B but is said to be highly specific for B:

DEFINITION 22. A **is highly specific for** B when the percentage of false positives is small.

DEFINITION 23. A **is completely screening for** B when there are no false negatives.

Hence A is completely screening for B if and only if $[\overline{A}B] = 0$, that is $c = 0$, from which it follows by Thm. 14 that,

THEOREM 33. A **is completely screening for** B **if and only if** B **is completely associated with** A.

If c is not zero but is relatively small, then A is not completely screening for B but is highly screening for B:

DEFINITION 24. A **is highly screening for** B when the percentage of false negatives is small.

It should be noted that a test which is highly specific for B may or may not have a large percentage of false negatives, and a test which is highly screening for B may or may not have a large percentage of false negatives. It follows from Thm. 30, 17 by Def. 21, 23 that,

THEOREM 34. A **is a completely reliable test for** B **when** A **is pathognomonic of and completely screening for** B.

A test which is easily satisfied is said to be a **sensitive** test; the more sensitive a test, the more likely that it will be satisfied, and the more likely it is that there will be false positives. When such a test is positive, the inference that B is present is less nearly certain, so that the positiveness of A is **weaker** evidence for the presence of B, and A is said to be a **weak** test for B. If, on the other hand, the test is less sensitive it may fail to be positive even when B is present, and there is a greater likelihood of false negatives. But when A **is** positive it is stronger evidence for the presence of B, and the test is more specific for B. These terms may be roughly summarized as follows:

(18)

sensitivity	% of false positives	% of false negatives	strength	specificity	screening
high	larger	smaller	weak	lower	higher
low	smaller	larger	strong	higher	lower

An illustration is furnished by the Wasserman test for syphilis, in which the reagents may be altered so as to make the test more or less sensitive. If the test is made more sensitive, then the greater likelihood of a positive even in the absence of syphilis, so that there is likely to be a greater percentage of false positives, and the test is less specific; but now the likelihood of false negatives is less, and the test is a better screening test. If, on the other hand, the test is made less sensitive, then the less the likelihood of a false positive and the greater the likelihood of a false negative; a positive test is now stronger evidence of syphilis, and the test is more specific, though it is less screening.

Positive association of a test with a disease is illustrated by the positive blood Kahn test and syphilis. Let us suppose, for instance, that in a given universe ten percent of persons have syphilis, that ten percent of all persons with syphilis have (falsely) negative Kahn tests, and that one percent of all non-syphilitics have (falsely) positive Kahn tests. Then the distribution is as follows:

(19)

	syphilis	no syphilis	
+ Kahn	90	9	99
— Kahn	10	891	901
	100	900	1,000

We can here see the important distinction between percentage of false positives (or false negatives) and the percentage of non-syphilitics who have false positives (or of syphilitics who have false negatives): the percentage of false positives is 9, whereas the percentage of non-syphilitics having false positives is only 1; the percentage of false negatives is 1.1, whereas the percentage of syphilitics having false negatives is 10. It is seen by (19) that $ad = 80{,}190$, $bc = 90$, $ad - bc > 0$, so that, by Thm. 6, a positive blood Kahn and syphilis are positively associated.

Dissociation is illustrated by the test for free gastric HCl and primary gastric carcinoma. Let us suppose that in a given universe 1% of persons have primary gastric carcinoma, that no person with primary gastric carcinoma has free gastric HCl, and that 10% of all persons without primary gastric carcinoma have no free gastric HCl; this is shown in the following array:

(20)		primary gastric Ca	no primary gastric Ca	
	free gastric HCl	0	891	891
	no free gastric HCl	10	99	109
		10	990	1,000

Since a = 0, it follows by Thm. 24 that free gastric HCl and primary gastric carcinoma are (in the present illustration) completely dissociated, so that they are, by Thm. 26, negatively associated. Now suppose we write array (20) in the form

(21)		primary gastric Ca	no primary gastric Ca	
	no free gastric HCl	10	99	109
	free gastric HCl	0	891	891
		10	990	1,000

which contains exactly the same data as does (20). In this second array it is c which is zero, and it follows by Thm. 15 that "no free gastric HCl" and primary gastric carcinoma are completely associated, and are, by Thm. 16, positively associated. We saw above that "free gastric HCl" and primary gastric carcinoma are negatively associated, which is in accord with Thm. 10.

When A, B are positively associated (ad—bc $>$ 0) it is seen by (10) that the coefficient of association between A, B is positive, and it is seen by (9) and (17) that it decreases as the percentage of false positives increases and as the percentage of false negatives increases; thus, the reliability of A as a test for B increases as the coefficient of association increases. When, on the other hand, A, B are negatively associated (ad—bc $<$ 0) it is seen by (10) that the cofficient of association between A, B is negative. If we compare (7) with (1) we see that the coefficient of association between A, $\overline{B}$ is obtained from that between A, B by interchanging **a** with b and c with d; and it is seen by (10) that this double interchange merely changes the sign but not the numerical (absolute) value of the coefficient of association. It is similarly shown that

(22)
$$C_{A,\overline{B}} = -C_{\overline{A},B} = -C_{A,\overline{B}} = C_{A,B}$$

Now if from the fact that A is positive we can infer that $\overline{B}$ is present, we can equally infer that B is absent, so that A has the same **absolute reliability** as a test for B that it has as a test for $\overline{B}$. It is seen also, by (17), that if we make the double interchange of **a** with b and c with d, then the percentage of false positives and of false negatives of A as a test for B is replaced by the percentage of false positives and of false negatives of A as a test for $\overline{B}$; consequently, we may say that,

THEOREM 35. **The absolute reliability of A as a test for B varies as the absolute value of the coefficient of association between A, B.**

We can illustrate this by computing the coefficient of association between a positive blood Kahn and syphilis for the distribution (19):

$$ad = 80{,}190\ ,\ bc = 90$$
$$ad-bc = 80{,}100\ ,\ ad+bc = 80{,}280$$
$$(a+b)(c+d) = 89{,}199$$

$$C = 50\left[\frac{80{,}100}{80{,}280} + \frac{80{,}100}{89{,}199}\right] = 94.8$$

so that the positive blood Kahn has a high degree of absolute reliability in the diagnosis of syphilis, on the assumptions underlying (19). For the distribution (21) we compute as follows:

$$ad = 8{,}910\ ,\ bc = 0$$
$$ad-bc = 8{,}910\ ,\ ad+bc = 8{,}910$$
$$(a+b)(c+d) = 97{,}119$$

$$C = 50\left[\frac{8{,}910}{8{,}910} + \frac{8{,}910}{97{,}119}\right] = 54.6$$

so that the test for free gastric HCl does not have a high absolute reliability in the diagnosis of primary gastric carcinoma, on the assumptions underlying (21). We see that the absolute reliability is considerable less in the second example than in the first, in spite of the fact that the two attributes are completely associated in the second example (cf. Thm. 15) but not in the first; this is explained by the fact that in the second example there is a high percentage (90%) of false positives, whereas in the first the percentage of false positives is only 9 and the percentage of false negatives is only 1.1.

Consider now the following two distributions:

(23)

	B	$\overline{B}$	
A	30	0	30
$\overline{A}$	30	40	70
	60	40	100

(24)

	B	$\overline{B}$	
A	60	0	60
$\overline{A}$	0	40	40
	60	40	100

In each case $ad-bc > 0$, so that A, B are positively associated. In (23), A is completely associated with B, though B is not completely associated with A, whereas in (24) they are completely associated with each other and are equivalent (cf. Def. 9). It is evident that A is more reliable as a test for B in (24) than in (23), since it is completely reliable in the former (cf. Def. 20) but not in the latter, inasmuch as there are no false positives and no false negatives in (24), whereas there are 43% of false negatives in (23). This is reflected in the corresponding coefficients of association $C_{A,B}$ which are 78.6 and 100 in (23), (24), respectively.

It can now be made clear why we have defined the coefficient of association as in (9), and prefer it to that of Yule and Kendall[3], whose coefficient of association (with the factor 100 supplied) is simply

$$\frac{ad-bc}{ad+bc}$$

which has the value 100 in each of (23) and (24), so that the greater reliability is not reflected in a greater coefficient of association.

As we have stated earlier, the **usefulness** of a test for some particular purpose may not be measured by its absolute reliability: a test may be more useful for some particular purpose than another test with greater absolute reliability, depending upon whether we wish the test to be more specific or more screening.

Applications to Prophylaxis and to Therapy.

The third application to medicine of the theory of statistical association is in the evaluation of a prophylactic measure A in the prevention of a disease B, or of a therapeutic measure A in the treatment of a disease B.

Suppose we wish to evaluate a vaccine in the prevention of a disease B; let A denote vaccination. It is clear that the vaccine cannot be tested unless it is used, nor can it be tested if it is given to all individuals in the universe, except if it is given to all and if all contract the disease, in which case it has been worthless; but even in this case we have not learned all we might have learned, because the vaccine may have been not only worthless but actually deleterious: if it had not been used there might have been some individuals who did **not** contract the disease. Hence in all cases we must have $[A] \neq 0$ and $[\overline{A}] \neq 0$. If no individual develops the disease, whether or not he was vaccinated, then we have no evidence as to the efficacy of A; only if all individuals develop the disease can we infer anything as to its efficacy, namely that it had no value in preventing B. Except for this last case, then, we must have $[B] \neq 0$ and $[\overline{B}] \neq 0$. And now it follows that we consider only cases in which

$$a+b > 0,\ c+d > 0,\ a+c > 0,\ b+d > 0$$

which is the same as (6) and (13). In the present application it is the c+d individuals in whom A was not carried out

who constitute the **controls.** The same considerations apply to the use of a therapeutic measure A in the treatment of a disease B, in which A denotes the use of the therapeutic measure and B denote some well-defined outcome of the disease, such as survival, cure, or something else.

CONCLUSION

The evaluation of a diagnostic criterion is **not possible without controls;** the evaluation of a prophylactic or therapeutic measure is **not possible without controls,** with the single exception of a case in which the measure is of no value whatever. The controls in the evaluation of a diagnostic criterion A for a disease B consist of those individuals without B, that is, of those individuals of type $\overline{B}$. The controls in the evaluation of a prophylactic or therapeutic measure A for the prevention or treatment of a disease B consist of those individuals in whom the measure was not employed, that is, of those individuals of type $\overline{A}$.

REFERENCES

1. Carlson, A. J.: "The Complex Causes of Alcohol Addiction." *Proceedings of the Insttute of Medicine of Chicago,* vol. 16, no. 8, Nov. 15, 1946.
2. Darwin, C.: *The Origin of Species,* 1859.
3. Yule, G. U. and Kendall, M. G.: *An Introduction to the Theory of Statistics,* p. 44. Charles Griffin & Co., Ltd., London, 1940.
4. Zeisler, E. B.: "Essays on Medical Statistics. I. Introduction to Medical Statistics." *The Chicago Medical School Quarterly, 19*:186, 1958 (Nov.)

ABSTRACTS SECTION

WEST, MICHAEL, M.D., and ZIMMERMAN, HYMAN J., M.D. Serum Enzymes in Disease. I. Lactic Dehydrogenase and Glutamic Oxalacetic Transaminase in Carcinoma.

From the Medical Service, West Side Veterans' Administration Hospital, the Departments of Medicine, University of Illinois College of Medicine, and The Chicago Medical School.

Serum lactic dehydrogenase (LD) was found to be elevated in 76 of 173 patients with carcinoma. Elevations were almost always associated with dissemination, usually to the liver. Only four patients without hepatic metastases exhibited elevated LD levels.

Serum gultamic oxalacetic transminase levels were found to be elevated in 43% of 121 patients. In all but one patient elevations were related to hepatic or myocardia metastases.

The majority of patients with hepatic metastases had abnormal levels of LD, GOT, alkaline phosphatese, or sulfobromophthalein (Bromosulphalein) (BSP) excretion. There was no demonstrable difference in the sensitivity of each of the 4 parameters in detecting metastases.

Analysis of the results suggested that LD elevations reflect the mass of tumor and GOT elevations, necrosis of tissue invaded by metastases.

* * * *

ZIMMERMAN, HYMAN J., M.D.; WEST, MICHAEL, M.D.; and HELLER, PAUL, M.D. Serum Enzymes in Disease. II. Lactic Dehydrogenase and Glutamic Oxalacetic Transaminase in Anemia.

From the Departments of Medicine, West Side Veterans' Administration Hospital, The Chicago Medical School, and the University of Illinois College of Medicine.

The serum levels of lactic dehydrogenase and glutamic oxalacetic transaminase have been studied in 82 patients with various types of anemia.

LD levels were consistently elevated in patients with megaloblastic anemia and sickle-cell anemia. GOT levels were elevated when there was other evidence of hepatic disease.

It is suggested that LD elevations in sickle-cell anemia are probably related to hemolysis and hepatic necrosis. In megaloblastic anemia this probably results from other factors.

* * * *

WEST, MICHAEL, M.D.; HELLER, PAUL, M.D.; and ZIMMERMAN, HYMAN J., M.D. Serum Enzymes in Disease. III. Lactic Dehydrogenase and Glutamic Oxaloacetic Transaminase in Patients with Leukemia and Lymphoma.

From the Departments of Medicine, West Side Veterans' Administration Hospital; University of Illinois College of Medicine; and The Chicago Medical School.

Serum lactic dehydrogenase was found to be consistently elevated in both untreated acute leukemia and untreated chronic granulocytic leukemia. Only 1 of 29 patients with chronic lymphocytic leukemia had an elevated level of lactic dehydrogenase.

Serum glutanic oxalacetic transaminase levels were infrequently elevated in all three types of leukemia. Elevations, when present, usually occurred terminally.

Approximately half of the patients with malignant lymphoma exhibited elevations of both serum enzymes. These elevations were associated with widespread dissemination and could be correlated with abnormal liver function.

Normal lactic dehydrogenase level in acute leukemia and chronic granulocytic leukemia appeared to be related to clinical remission.

* * * *

WEST, MICHAEL, M.D. and ZIMMERMAN, HYMAN J., M.D. Serum Enzymes in Disease. IV. Lactic Dehydrogenase and Glutamic Oxalacetic Transminase Levels in Renal Disease.

Departments of Medicine, West Side Veterans' Administration Hospital, The Chicago Medical School, and the University of Illinois.

Forty-three of 71 patients with renal disease were found to have elevated lactic dehydrogenase levels in the serum. Although azotemia was common in this group, there was no direct relationship between the serum enzyme levels and the degree of azotemia. The degree of anemia, acidosis, proteinuria, and hypercholesterolemia could not be related to the serum lactic dehydrogenase levels. There was a statistically significant (P— 0.05) inverse correlation between the serum albumin levels and the lactic dehydrogenase levels.

Experimental renal failure produced in rabbits by uretral ligation or by dehydration resulted in an inconsistent occurrence of elevated serum lactic dehydrogenase.

The serum glutamic oxalacetic transaminase levels were found to be elevated in only 6 patients of 63 with renal disease. Four of these 6 had coincidental hepatic disease adequate to account for the elevated glutamic oxalacetic transaminase levels. Another had delirium tremens and acute renal failure. In one patient with renal failure, but no demonstrable hepatic disease, this enzyme was also found to be elevated.

The factor(s) responsible for the elevated serum lactic dehydrogenase levels in patients with renal disease could not be elucidated.